THE PHILOSOPHY OF ECONOMICS

AN ANTHOLOGY

Third Edition

Edited by
Daniel M. Hausman

CAMBRIDGE

The Philosophy of Economics

An Anthology

Third Edition

Edited by

DANIEL M. HAUSMAN

University of Wisconsin–Madison

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521883504

First published in print format 2007

ISBN-13 978-0-511-37141-7 eBook (NetLibrary)
ISBN-10 0-511-37141-1 eBook (NetLibrary)

ISBN-13 978-0-521-88350-4 hardback
ISBN-10 0-521-88350-4 hardback

Contents

Introduction

Premises assumed without evidence, or in spite of it; and conclusions drawn from them so logically, that they must necessarily be erroneous.

– Thomas Love Peacock, *Crochet Castle*

Ever since its eighteenth-century inception, the science of economics has been methodologically controversial. Even during the first half of the nineteenth century, when economics enjoyed great prestige, there were skeptics like Peacock. For economics is a peculiar science. Many of its premises are platitudes such as "Individuals can rank alternatives" or "Individuals choose what they most prefer." Other premises are simplifications such as "Commodities are infinitely divisible," or "Individuals have perfect information." On such platitudes and simplifications, such "premises assumed without evidence, or in spite of it," economists have erected a mathematically sophistical theoretical edifice, whose conclusions, although certainly not "necessarily erroneous," are nevertheless often off the mark. Yet businesses, unions, and governments employ thousands of economists and rely on them to estimate the consequences of policies. Is economics a science or isn't it?

This is a complicated question. What does it mean to assert or deny that economics is a science? To be called a science is, no doubt, an honor. As the scientific credentials of economists rise, so do consulting fees. But what question is one posing when one asks, "Is economics a science?" Is one inquiring about the goals of economics, about the methods it employs, about the conceptual structure of economic theory, or about whether economics can be reduced to physics? If economics is a science, is it the same *kind* of science as are the natural sciences?

During the last generation, interest in philosophical questions concerning economics has increased enormously. Twenty-five years ago, when I was working on the first edition of this anthology, this interest was already growing, with philosophers, economists, other social scientists, and ordinary citizens all showing more curiosity about what sort of an intellectual discipline economics is and what sort of credence its claims merit. At the time, many turned to the literature on methodology because of doubts about the value of economics. After the economic successes of the generation following World War II, economic growth stalled in the 1970s, and many came to doubt that *anybody* knew how to restore prosperity without rekindling inflation.

A decade later, at the time of the second edition, things looked brighter for economics, although there were still doubts about how to restore prosperity without aggravating budget deficits, how to reinstitute markets in state-controlled economies without precipitating economic collapse, and how to alleviate widespread misery in the so-called developing countries. In that atmosphere, it is not surprising that economists turned to methodological reflection in the hope of finding some flaw in previous economic study or, more positively, some new methodological directive to improve their work. Nor is it surprising that ordinary citizens, whose opinions of economists are more influenced by the state of the economy than by systematic evaluation of economic theories, should wonder whether there might be something awry with the discipline.

Today, in 2007, in contrast, economists are riding high. Although there have been serious economic problems during past fifteen years, such as the international financial crisis in 1997, continued high unemployment in Europe, and a prolonged and severe recession in Japan, nevertheless, there has been significant economic growth in developed economies, which have generally prospered. Serious problems remain in the formerly socialist countries, but conditions have stabilized and for the most part improved. And rapid economic growth in the two most populous countries on earth, India and especially China, has transformed the economic landscape. Although it is overly optimistic to claim that the central economic problems have been solved (especially in the light of the disastrous performance of the economies of many of the poorest countries in the world), such a claim today, unlike a generation ago, would not strike most people as absurd.

While the doubts about the value of economics that helped fuel the interest in economic methodology that began in the 1970s have receded, the theoretical reasons to be interested in economic methodology have only grown stronger. In previous editions, I identified three theoretical reasons. First, not only economists but also anthropologists, political scientists, social psychologists, and sociologists influenced by economists have argued that the "economic approach" is the only sensible theoretical approach to the study of human behavior. This provocative claim – that economics is the model that *all* social sciences must follow – obviously makes methodological questions concerning economics more important to other social scientists.

In the 1970s and 1980s, it was ironic that some economists were making grandiose claims for the universal validity of the economic approach to human behavior at the same time that others had serious qualms about their own discipline. As those qualms have faded, so has this irony. There is, however, a second ironical twist, which constitutes the second theoretical reason why interest in the methodology of economics has increased. During the same period that grand claims have been made for the economic approach to human behavior, cognitive psychologists and economists impressed by the work of cognitive psychologists have shown that many of the fundamental claims of modern mainstream economics are refuted by economic experimentation. The rapid expansion of experimentation, which is discussed in Vernon Smith's essay (Chapter 18) and of behavioral and neuroeconomics, which is discussed in Colin F. Camerer's essay (Chapter 19), raise intriguing methodological questions.

Finally, there are special reasons why philosophers have become more interested in the methodology of economics. Contemporary philosophers of science have become convinced that a great deal can be learned about how science ought to be done from studying how science actually is done. Although most philosophers who are interested in the sciences study the natural sciences, economics is of particular philosophical interest. Not only does it possess the methodological peculiarities sketched above, but moral philosophers, whether attracted or repelled by the tools provided by economists and game theorists, need to come to terms with welfare economics (which is discussed in Part III of this anthology).

For these reasons, it is not surprising that there is so much interest in the methodology of economics. At the same time that triumphant economists are claiming to have found the one true path for all the social sciences, psychologists, behavioral economists, and neuroeconomists are challenging the basic generalizations of economics and arguing for a different way of doing economics. Philosophers of science are at the same time turning their attention to the peculiarities of particular disciplines, such as economics. The renewed interest in economic methodology over the last generation comes after decades during which the subject was largely ignored by philosophers, while the philosophical efforts of economists – in many cases prominent ones – were sporadic and often polemical.

This volume aims to assist those interested in the methodology of economics by providing a comprehensive and up-to-date introduction to the subject. My hope is that this book will be useful both as a research resource and as a teaching tool. It provides an introduction to a wide range of methodological issues and and to a wide range of positions which have been taken with respect to these issues.

Unlike a textbook, this anthology also provides some historical perspective. Methodological questions concerning economics – questions about the goals of economics, the ways in which economic claims are established, the concepts of economics and their relation to concepts in the natural sciences and so forth – are all philosophical questions, and in philosophy it is generally a mistake to ignore the works of the past. Past wisdom cannot be encapsulated in a textbook, and original works cannot be consigned to intellectual historians. Much of what a philosophical text has to teach lies in its relationship to its intellectual context and in the nuances of its argumentative turns. There is, I believe, a great deal to be learned about economic methodology from studying directly how intellectual giants like John Stuart Mill or Karl Marx dealt with the problems. Those who wish to think seriously about the methodology of economics should know its history, too.

Some introductory material may help the reader to understand the essays reprinted here. At the beginning of each part, I offer a few comments about its contents. The remainder of this general introduction provides general background to make the various essays more accessible. Capsule introductions to the philosophy of science, to economic theory, and to the history and contemporary directions of work on economic methodology follow.

An Introduction to Philosophy of Science

As science is one sort of human cognitive enterprise, so philosophy of science is a part of epistemology (the theory of knowledge), although philosophers of science also face questions concerning logic, metaphysics and even ethics and aesthetics. One can find discussions of issues in the philosophy of science in the works of pre-Socratic philosophers, but philosophy of science as a recognizable subspecialty only emerged during the nineteenth century. Important names in the early development of modern philosophy of science are David Hume and Immanuel Kant in the eighteenth century, and John Stuart Mill and William Whewell in the nineteenth century. At the end of the nineteenth century, philosophy of science emerges as a subdiscipline with monographs mainly by scientists or historians of science such as Ernst Mach, Pierre Duhem and Henri Poincaré. In the first half of the twentieth century, the so-called logical positivists (many of whom also had backgrounds in science) dominated thinking about the philosophy of science, although Karl Popper's views also were influential. Contemporary philosophy of science is a lively area of research and controversy. Although there is considerable agreement about fundamentals, the details concerning matters such as explanation or confirmation are hotly contested. There is no standard doctrine or detailed orthodoxy.

The issues with which the philosophy of science has been concerned that are most relevant to economics can be divided into five groups:

1. *Goals* What are the goals of science and of scientific theorizing? Is science primarily a practical activity that aims to discover useful generalizations, or should science seek explanations and truth?
2. *Explanation* What is a scientific explanation?
3. *Theories* What are theories, models, and laws? How are they related to one another? How are they discovered or constructed?
4. *Testing, induction and demarcation* How does one test and confirm or disconfirm scientific theories, models and laws? What are the differences between the attitudes and practices of scientists and those of members of other disciplines?
5. Are the answers to these four questions the same for all sciences at all times? Can human actions and institutions be studied in the same way that one studies nature?

This grouping of the questions with which philosophers of science have been concerned is intended only to help organize the discussion that follows. I have omitted issues concerning the unobservable postulates of scientific theories, which were of great importance to the logical positivists and their immediate successors, because they are less important to economics.

Contemporary philosophy of science is best understood against the background of positivist and Popperian philosophy of science, which are still influential among economists. So in discussing the questions listed here, I shall spend some time talking about the positivist and Popperian ancestors of contemporary views.

The Goals of Science

There are two main schools of thought. *Scientific realists* hold that in addition to helping people to make accurate predictions, science should *also* discover new truths about the world and explain phenomena. The goal is truth, and enough evidence justifies claims to have found the truth, although realists recognize that the findings of science are subject to revision and correction with the growth and improvement of science. *Antirealists* may be *instrumentalists*, who regard the goals of science as exclusively practical, or antirealists may instead disagree with realists mainly about whether the unobservables postulated by scientific theories exist, whether claims about them are true or false, and whether observable evidence can establish claims about unobservables. Notice that instrumentalists do not repudiate theorizing. They agree with realists that theories are important. But they locate their importance exclusively in their role in helping people to anticipate and control phenomena. In his influential essay, "The Methodology of Positive Economics" reprinted in this anthology, Milton Friedman espouses a narrowly instrumentalist view of science.

Who is right, realists or antirealists? There is no settled opinion among philosophers, and the fortunes of realism and instrumentalism have oscillated over the past few decades.[1] Scientists themselves are divided. Realism has a firm foothold in many areas (how many people doubt that DNA exists or that it carries a genetic code?), but the problems and peculiarities of quantum mechanics have led many physicists to a modest view of the goals of science and to an antirealist view of claims about quantum phenomena. For a discussion of the relevance of realism versus antirealism to economics, see Uskali Mäki's and Tony Lawson's essays in Part V.

Someone who hopes that science can discover new truths about the world through its theorizing need not find theories *valueless* unless they are true. Ptolemy's astronomy, which places the earth in the center of the solar system, was used for navigational purposes for centuries after it was refuted. There is no reason why a realist cannot use Ptolemy's theory to navigate. The realist wants more from science than such merely useful theories, but that is no reason to throw away something that works.

Scientific Explanation

Explanations answer "Why?" questions. They remove puzzlement and provide understanding. Often people think of explanations as a way of making unfamiliar phenomena familiar, but in fact explanations often talk of things that are much *less* familiar than what they seek to explain. What could be more familiar than that water is a liquid at room temperature? Certainly not the explanation physicists give for its liquidity.

Philosophers disagree about what is central to a scientific explanation. Logical positivists and their logical empiricist successors took scientific explanations to show that the event or regularity to be explained follows from a deeper regularity. A scientific explanation shows us that what is to be explained could have been expected to happen. This notion of explanation goes back to the Greeks, but it receives its best systematic development in the twentieth century in essays by Carl Hempel.[2] Hempel develops two main models of scientific explanation, the deductive-nomological and the inductive-statistical models. The latter, as its name suggests, is concerned with probabilistic explanations and attempts to extend the basic intuition of the deductive-nomological (D-N) model.

In a deductive-nomological explanation, a statement of what is to be explained is *deduced* from a set of *true* statements which includes *essentially* at least one *law*. Schematically, one has:

True statements of initial conditions
Laws
———————————————
Statement of what is to be explained

The line represents a deductive inference. One deduces a description of an event or regularity from laws and other true statements. It is essential that there be at least one law. To deduce that this apple is red from the true generalization that all apples in Bill's basket are red and the true statement that this apple is in Bill's basket does not explain why the apple is red. "Accidental generalizations," unlike laws, are not explanatory.

The D-N model is an account of deterministic, or nonstatistical explanations. If one has only a statistical regularity, then one will not be able to *deduce* what is to be explained, but one may be able to show that it is highly probable, which is what Hempel's inductive-statistical model requires.

Even when limited to nonstatistical explanations, the D-N model faces counterexamples. An argument may satisfy all the conditions of the D-N model without being an explanation. For example, the fact that someone takes birth control pills regularly does not explain why they do not get pregnant, if the person never has intercourse or is a male. But not getting pregnant is all the same an implication of the "law" that those who take birth control pills as directed do not get pregnant.[3] One can deduce the height of a flagpole from the length of its shadow, the angle of elevation of the sun, and the law that light travels in straight lines, but doing so does not explain the height of the flagpole. A similar deduction does, however, explain the length of the shadow.[4]

What has gone wrong? The intuitive answer is that taking birth control pills has no causal influence on whether a woman who never has intercourse gets pregnant, and men cannot get pregnant whether or not they take birth control pills. Similarly, sunlight and shadow have no significant causal influence on the height of flagpoles. It seems that explanations of events and states of affairs typically cite their *causes*.[5] There are, however, two problems with "explanations cite causes" as a theory of explanation. First, although most explanations of events and states of affairs are causal explanations, not all are. Second, saying that explanations cite causes is not by itself very informative. Without a theory of causation, a causal theory of explanation is empty, and even with a theory of causation, it only scratches the surface to maintain that to explain is to cite a cause. The existence of the sun is causally relevant to the wheat harvest, but it does nothing to explain the price of wheat.

The explanation of human behavior introduces special difficulties. Most explanations of human action take a simple form. One explains why an agent purchased some stocks or changed jobs by citing relevant beliefs and desires of the agent. When economists explain behavior in terms of utility functions, they offer explanations of just this kind.

This familiar kind of explanation is philosophically problematic. If one attempts to construe such explanations as elliptical or sketchy deductive-nomological explanations, one finds that it is hard to find any substantial and plausible laws implicit in them. What apparently do the explaining are platitudes such as "People do what they most prefer." Some philosophers have argued that generalizations like these are not empirical generalizations at all. They are instead implicit in the very concepts of action and preference.[6] According to these philosophers, explanations of human behavior differ decisively from explanations in the natural sciences. In explaining why someone did what he or she did, one does not subsume their action under some general regularity. Instead, one gives the agent's *reasons*.

It is true that in explaining an action one gives the agent's reasons for performing it. But do explanations in terms of reasons differ fundamentally from explanations in the natural sciences? Can they be seen as (roughly) deductive-nomological or as causal? Can they be assessed in the same way that explanations in the natural science are assessed? Philosophers disagree on these questions. Most writers on economics have attempted to assimilate explanations in economics to explanations in the natural sciences. Why cannot explanations in terms of reasons *also* be scientific explanations in terms of causes?[7] But there is a considerable minority, which includes distinguished economists such as Frank Knight (Chapter 4), who have argued that explanations of actions in terms of the reasons for the actions differ in some fundamental way from ordinary scientific explanations.

Scientific Theories and Laws

Most philosophers have argued that science proceeds by the discovery of theories and of laws, but economists are more comfortable talking about *models* than about laws and theories. Over the last two decades, philosophers have begun to catch up,[8] and there is a new philosophical literature that permits a more satisfactory characterization of theorizing in economics.

Economists do sometimes talk in terms of laws. They speak of the law of demand, Say's Law, the law of one price, and so forth. So let us begin with some words concerning laws and the role they play in science. The laws of sciences are not, of course, prescriptive laws dictating how things *ought* to be. (It is not as if the Moon would like to leave its orbit around the earth, but is forbidden to do so by a gravitational edict.) Scientific laws are instead (speaking roughly) regularities in nature. But they are not just regularities. Consider the generalization, "No gold nugget weighs more than 1,000 tons." Even if it is true everywhere and for all time, this generalization appears to be merely "accidental" and of no explanatory value. What then is the difference between an accidental regularity and a genuine law?

Rather than canvas the unsatisfactory answers philosophers have considered, let us step back and ask whether, however the analysis comes out, economics has any genuine laws. Consider, for example, the law of demand. It says, roughly, that when the price of something goes down, people seek to buy more of it, and when the price goes up, people want to buy less. Unlike physical laws such as Boyle's law, which states that the pressure and volume of a gas are inversely proportional, the "law" of demand is asymmetrical: it links causes (price changes) to effects (changes in demand). If an increase in demand comes first, the price will go up rather than down. Second, the "law" of demand is (at least when stated this way) not a universal truth. For example, if there is a change in tastes at the same time that the price drops, demand might not increase. So perhaps the concept of a law is not a useful one for those interested in economic methodology.

The issues here are complicated, because of the possibility of subtle reformulations of claims such as the "law" of demand. One might, for example, argue that such laws carry *ceteris paribus* qualifications: other things being equal, price increases lessen demand and price decreases increase demand. In my own work, I have defended this idea, which goes back to John Stuart Mill (the first selection in this volume). So I do not think that this project is misconceived. According to the deductive-nomological model of explanation, economists can use generalizations such as the law of demand to explain economic phenomena only if those generalizations are genuinely laws.

Nevertheless, there is a good deal to be said for adopting an explicitly causal view of explanation such as James Woodward's, which does not depend on citing any laws. Whether or not the law of demand is truly a law, there are specific domains in which the generalization is nearly always true and in which one can rely on it to pick out the causes of price changes.

The other intellectual constructs emphasized by the logical empiricists, scientific theories, also do not fit economics very well. One of the features the positivists took to be crucial to theorizing – the postulation of unobservable entities and properties to explain observable phenomena – is unusual in economics. (Even though beliefs and preferences are apparently unobservable, they are obviously not new postulations of economists.) More importantly, when economists talk about theories, they usually talk about branches of economics (such as game theory, or the theory of the firm, or the theory of monopolistic competition) rather than anything analogous to Newton's theory of gravitation or Maxwell's theory of electromagnetic radiation.

Theories in the natural sciences appear to be collections of lawlike statements that "work together" to help describe, predict, and explain phenomena in some domain. The logical positivists made the notion of "working together" precise, by arguing that theories form deductive systems. According to the positivists, theories are primarily "syntactic" objects, whose terms and claims are interpreted by means of "correspondence" rules.[9] Let me explain.

Influenced as they were by the dramatic breakthroughs in formal logic at the end of the nineteenth and the beginning of the twentieth century, the logical positivists conceived of deducibility as a *formal* relationship between sentences, which is independent of the *meaning* of the sentences. For example, one can infer the sentence "*r*" from the sentence "*s* and *r*" without knowing anything about what the sentences "*s*" or "*r*" assert. Logicians explored the possibility of constructing formal languages in which the ambiguities of ordinary languages would be eliminated. In these formal languages, there would be a sharp separation between questions concerning syntax and semantic questions concerning meaning and truth.

The logical positivists hoped to be able to express scientific theories in formal languages. From the axioms of the theory, all theorems would follow purely formally (just as "*r*" follows from "*s* and *r*"). For the theory to have meaning and to tell us about the world, it would still need an interpretation. "Correspondence rules" were supposed to provide that interpretation and to permit theories to be tested. Originally, correspondence rules were conceived of as explicit definitions for each of the theoretical terms, but the positivists soon realized that the relationship between theory and observation is more intricate.

Scientific theories cannot usually be formalized in the way in which the logical positivists hoped, and the positivist view of theories does not do justice to the way in which theories are constructed or used. Furthermore, the problems of relating theory to observation, in the form in which the positivists posed them, are intractable, and problems about characterizing lawlike statements remain. Many philosophers of science now settle for a looser informal construal of theories as collections of interpreted lawlike *statements* rather than uninterpreted, purely syntactic sentences, which are systematically related to one another.

The really pressing philosophical task for those interested in economics is to come up with an understanding of scientific *models*, because economic theorizing relies mainly on models. Models in the sciences, unlike theories, may be material (like the scale models of airplanes tested in wind tunnels) as well as linguistic; however, like laws and theories, they are representational. Unlike laws and some theories, models are manipulated, explored, and modified. Although it is sometimes appropriate to ask whether parts of models are true or false, economists more often assess models in terms of their fruitfulness or usefulness.

One view of models, which I have defended (and which is criticized in the essay by Sugden, reprinted as Chapter 26), takes them to be of the same logical type as are predicates such as "has two legs," or definitions of such predicates.[10] According to this view, a model of consumer choice among two commodities does not make assertions about the world. It is instead a predicate such as "is a two-commodity consumption system" or a definition of such a predicate. Of course, economists do make claims about the world. They do so by *using* models, by asserting that the predicates that models constitute or define are true or false of systems of things in the world.

Drastically oversimplifying this view, it maintains that instead of offering "theories" like "All bodies attract one another with a gravitational force," scientists offer "models" like "Something is Newtonian system if and only if all bodies in it attract one another with a gravitational force and . . . ," and that scientists then use such 'models' to make empirical claims such as "The universe is a Newtonian system." Given this parody, one might wonder why serious philosophers defend the predicate view of models.

There are two reasons. First, if one hopes to be able to reconstruct the claims of science formally, the predicate view has significant technical advantages. Second, the predicate view offers a useful way to schematize the *two* kinds of achievements involved in constructing a scientific theory. Although what ultimately count are the claims that models permit scientists to make about the world, science does not proceed by spotting correlations among already known properties of things. An absolutely crucial part of the scientific endeavor is the construction of new concepts, of new ways of classifying phenomena. And much of science is devoted to thinking about these concepts, relating them to other concepts and exploring their implications. This kind of endeavor is prominent in economics, where economists often explore the implications of perfect rationality, perfect information and perfect competition, without immediate concerns about empirical application or testing.

Assessment and Demarcation

Most people are empiricists about theory assessment: they believe that the evidence that ultimately leads scientists to accept or to reject claims about the world should be perceptual or observational evidence. According to empiricists, economists should believe that individuals generally prefer more commodities to fewer, if and only if this claim is borne out by experience.

Empiricism is not completely uncontroversial. Kant argued in his *Critique of Pure Reason* that there are some "synthetic" truths about the world such as the axioms of Euclidean geometry that can be known "*a priori*" – that is, without specific sensory confirmation. He maintains that these propositions are implied by the very possibility of having any conscious experience of the world. No specific observations or experiences could ever lead us to believe that such propositions were false.

Modern physics has not dealt kindly with Kant's view that the axioms of Euclidean geometry are *a priori* truths, but the Kantian view that there are synthetic *a priori* truths still has supporters among so-called Austrian economists, especially Ludwig von Mises and his followers. They argue that the fundamental postulates of economics are synthetic *a priori* truths.[11] I shall not discuss the Austrians' epistemological views, but the reader should be aware that some methodologists question empiricist views on assessment.

Despite their "obviousness," empiricist views of the assessment of claims about the world encounter serious problems. First, it seems implausible to claim that definitional truths such as "Triangles have three angles" require testing or that our confidence in such claims rests on the results of observations. Nor do we need experiments to know that a claim such as "This square is circular" is false. The logical positivists responded by distinguishing synthetic claims – claims about the world – from analytic or contradictory claims whose truth or falsity depend solely on logic and on the meanings of the terms in such claims.[12]

Even confining oneself to synthetic claims, serious problems remain. As Hume argued in the eighteenth century, observation only establishes the truth of singular statements about particular events or about properties of things at particular times and places. On what, then, is our confidence in generalizations or in singular statements about instances not yet observed based? As Hume put it:

> If a body of like color and consistency with that bread which we have formerly eaten be presented to us, we make no scruple of repeating the experiment and foresee with certainty like nourishment and support. Now this is a process of mind or thought of which I would willingly know the foundation.[13]

In other words, Hume is issuing a challenge: Show me a good argument whose conclusion is some generalization or some claim about something not observed and whose premises include only reports of sensory experiences. Such an argument cannot be a deductive argument, because such inferences are fallible: the next slice of bread might be fatal. Nor will an "inductive" argument do, as we have only inductive and thus question-begging grounds to believe that such arguments are good ones.

This is Hume's *problem of induction*. It is primarily a problem concerning how singular claims about unobserved things or generalizations are to be *supported* or *justified*. It is not mainly a problem about the discovery of generalizations. In my opinion, Hume's problem of induction is, as stated, insoluble.

If this problem of induction cannot be solved, there are two options. One is to deny that there are ever good reasons to believe generalizations about the world, no matter how much purported evidence one has. This is the skeptical conclusion Hume drew – although he confessed that when he left his study he could not act on it. Alternatively, one can criticize Hume's description of the problem. I prefer the latter course. What is wrong with Hume's problem of induction is Hume's view of what justification demands. Hume wants a separate argument for every generalization with only reports of sensory experiences as premises. If instead one relaxes the demands on justification and one permits the premises in justificatory arguments to include all of our purported scientific knowledge about the world, then one faces the difficult but not impossible problems of inductive inference that scientists actually grapple with. Observations and experiments play a crucial role in the expansion and correction of empirical knowledge, but people need not trace their knowledge claims back to an experiential foundation.[14] To borrow a metaphor, learning about the world is like rebuilding a ship while staying afloat in it. In learning more about the world, people rely both on observation and on the vast body of knowledge that they think they already have.

The ship metaphor is due to Otto Neurath, who was a member of the Vienna Circle, the main wellspring of logical positivism. Yet the logical positivists did not for the most part endorse such a holistic view of scientific knowledge. Instead considerable efforts were made by Rudolf Carnap and others to develop an inductive logic, a canon of thought whereby conclusions could be established with a certain probability from premises that included only basic logic and mathematics and reports of observations.[15] These efforts were not successful, but Carnap's work helped lead to more promising modern approaches.[16]

Karl Popper's views on induction are more radical. Popper recognized in the 1930s that the results of experiments and observations bear on the truth or falsity of claims about the world only within the context of a body of tentatively accepted beliefs.[17] But he then introduced a further twist. He argued that generalizations such as "All copper conducts electricity" can be *falsified* by singular statements reporting the results of observations, even though they cannot be *verified*. In fact, Popper argued that there is no such thing as confirmation! (He says, instead, that scientific generalizations may be "corroborated," but he maintains that corroboration provides no grounds to believe that a theory is correct or a reliable basis for prediction.) Generalizations remain no more than tentative conjectures, no matter how often we fail to falsify them.

Many have read Popper as suggesting that generalizations can sometimes be conclusively proven to be false on established premises which include only reports of observations. The problem of induction is thus "solved" by accepting half of Hume's skeptical conclusion: There are never good reasons to believe that universal generalizations are true. What saves us from skepticism and generates scientific progress is the possibility of finding good reasons to believe that generalizations are false. Science proceeds by making bold conjectures and eliminating errors.

Popper explicitly disavowed this simple interpretation of his position.[18] In his view, reports of observations are fallible and open to revision. As a matter of convention one accepts them as true in the course of testing a generalization. In doing so, one is taking an unavoidable risk of rejecting the generalization, even though it is true. Moreover, one can rarely infer the falsity of interesting claims in science merely from singular observation reports. For example, to use observations of choices in the economics laboratory to test game theory, one has to make assumptions concerning what factors influence preferences. In testing a theory, scientists deduce an implication from that theory, *conjoined with* subsidiary hypotheses and statements of initial conditions. If the implication is not borne out by observation, scientists must take risks and decide that the problem lies in the particular theory being tested, not in the unavoidable additional premises.

In autobiographical comments, Popper maintains that what drove him into the philosophy of science was what he calls "the problem of demarcation": What is the difference between a scientific theory and a theory which is not scientific?[19] Although formulated differently, this was a driving question for the logical positivists, too. They wanted to be able to distinguish scientific theories from "meaningless" metaphysics and to contribute to the further development of science. As stated earlier, the problem of demarcation concerns the distinction between scientific *theories* and other sorts of theories. But Popper is often concerned instead to distinguish those *attitudes, rules and practices* that distinguish a scientific community from other attitudes and practices. What matters is often not the theory, but what people think of it and what they do with it. Newton's theory of motion could become the dogma of some strange sect, while, in contrast, astrology can be subjected to scientific scrutiny. The more important problem of demarcation concerns the difference between the attitudes of scientists and nonscientists, not the difference between scientific theories and other sorts of theories.

According to Popper, what is special about scientists is that they have a "critical attitude." They follow methodological rules directing them to make bold conjectures and then seek out the harshest possible tests of them. These rules require that when the conjectures fail those tests, scientists do not make excuses. Instead they should regard the theories as refuted, and they should then propose and scrutinize new conjectures.[20] As many have noted, including Thomas Kuhn and Imre Lakatos, it is a good thing that scientists do not follow these rules.[21] Because theories always face unresolved difficulties, these rules demand that they all be rejected. But theories are too important to the practice of science to be surrendered until alternatives are available. And alternatives are not easily generated.

The questions Popper asks may be more important than the answers he argues for. Successors such as Kuhn and Lakatos and a number of sociologists of science have followed Popper in attempting to clarify what sort of disciplines the sciences are. Yet current investigations of assessment and demarcation differ not only from the positivists' efforts, but from Popper's as well. As completely opposed as the Popperian and positivist approaches were, both conceived of theory assessment in terms of the confrontation of single theories with data. Most contemporary philosophers of science reject this way of approaching the problems. Instead of thinking about the problems of theory *assessment*, they are concerned with the problems of theory *comparison* and *choice*. Testing is a many-sided confrontation among alternative theories and data. Furthermore, there are many choices to be made among theories, not just one. A scientist may for example reasonably believe that theory T is better confirmed than theory T', but that T' offers more interesting research possibilities. Although most contemporary philosophers of science agree that there are many different problems of theory assessment and that one must address them in terms of *choices* among alternatives, disagreements remain about what conclusions to draw.

One view, which many attribute to Thomas Kuhn, is to question whether theory choices are rationally defensible. In his classic *Structure of Scientific Revolutions*, Kuhn offers a view of science and of philosophy of science that differed sharply from the logical empiricist orthodoxy at the time he was writing. With the help of vivid examples from the history of science, Kuhn emphasizes how extensive are the constraints on ordinary scientific research. To determine the magnitude of a particular constant or to solve a detailed theoretical problem takes resources and energy, which scientists will not be willing to expend unless they are convinced that the general theoretical framework ("paradigm") within which they are working is more or less correct. Without such commitments, detailed esoteric research efforts would not be undertaken. Although the workaday, perhaps even dogmatic "normal science" that results does not aim at discovering novelties, it nevertheless, Kuhn argues, uncovers "anomalies" – problems that resist solution within the particular normal scientific tradition. Such anomalies can undercut the scientific community's confidence in the accepted paradigm and, given the construction of an alternative paradigm, can lead to a scientific revolution.

Kuhn's view of scientific revolutions is especially controversial. He seems to argue that disagreements in scientific revolutions can be so pervasive that no rational choice can be made.[22] Because scientists in different camps will have distinct views about standards of theory assessment and about how to conceive of the subject matter and practice of the science, consensus can be reached only through nonrational persuasion. According to this irrationalist interpretation of Kuhn, the paradigm that triumphs in a scientific revolution need not be objectively "better" than the paradigm it replaced.

Kuhn disavows such an extreme interpretation of his views, and many historians and most philosophers of science have found such irrationalist conclusions to be unjustified. Yet they live on in the work of some sociologists of science, who have defended even more extreme views. Some go so far as to deny that the phenomena that scientists study have any influence at all on the views that scientists defend.[23]

Yet in rejecting Kuhn's apparent irrationalism, one can still recognize the significance of his contribution to contemporary philosophy of science. Not only did Kuhn make philosophers aware of the complexity of scientists' commitments, but he did as much as anyone to convince philosophers that theorizing about science without careful attention to scientific practice was likely to be misleading. Even though few philosophers of science regard themselves as Kuhnians, most follow Kuhn on these points. Although Popper and many of the logical positivists were scientifically literate and intensely interested in the sciences, including particularly physics, contemporary philosophers of science tend to address problems in the philosophy of science at a lower level of abstraction and with greater attention to the details of scientific practice. Just as economists can only offer advice to a firm if they have learned what in fact makes firms run well, so philosophers can only offer advice to scientists if they have learned what in fact makes for good science. And, in my view, there is in general no way to learn about firms or science without studying firms or scientists.

A number of prominent philosophers of science have developed accounts of theory evaluation that recognize the complexities of scientific work without denying the rationality of science. Many approaches merit discussion, especially the work of the modern "Bayesians," but this introduction is not long enough to discuss them.

Something must, however, be said about Imre Lakatos's "Methodology of Scientific Research Programmes," which had a considerable influence on economic methodology in the 1970s and 1980s. Lakatos began his work on the philosophy of science as a follower of Popper. Although critical of many details, including Popper's view that scientific honesty demands an immediate readiness to surrender one's theory in the fact of an apparent disconfirmation, Lakatos insists that Popper's basic point remains valid: if scientists make empty excuses for their theories when they run into apparent difficulties, then they will never learn from experience. What philosophy of science should be concerned with, according to Lakatos, are not rules for assessing theories, but rules for *modifying* and *comparing* theories. Rather than asking, "Is theory *T* well or poorly supported by the data?" scientists want to know whether a new version of *T* is an improvement over the old. The central question concerning assessment is whether the proponents of *T* are making as much progress improving it as are the proponents of competing theories.

According to Lakatos, a modification of a theory is an improvement if it is not *ad hoc*. Modifications may be *ad hoc* in three ways.[24] If a modification of a theory has no new testable implications at all, it is empty and unscientific. Modifications that are not *ad hoc* in this first sense are "theoretically progressive." If the testable implications of theoretically progressive modifications are not confirmed by observation, then these modifications are not "empirically progressive," and they are thereby *ad hoc* in the second sense. Lakatos maintains that an extended process of theory modification is progressive overall, if the modifications are uniformly theoretically progressive and intermittently empirically progressive. As scientists revise their theories in the hope of improving them, the changes must always have new testable implications; and those testable implications must sometimes be borne out by experiment and observation. In addition, there must be continuity throughout this history of repeated modification. Economists do not make theoretical progress by tacking on unrelated generalizations from chemistry. Adding the generalization that copper conducts electricity to monetary theory results in new testable implications, but such a modification is *ad hoc* in a third sense.

Lakatos insists that science is and should be dominated by scientific research programs. These consist of a series of related theories that possesses a "hard core," which the "negative heuristic" insists must be preserved through all modifications of particular theories within the research program. In addition, the research program contains a "a positive heuristic" that directs scientists in making modifications. Particular changes within a research program should be assessed by considering to what extent they are theoretically and empirically progressive and to what extent they are in accordance with the positive heuristic of the research program. Competing research programs should be compared by examining their overall progressiveness. In Lakatos's view (in contrast to Kuhn's), science suffers when a single research program becomes dominant.

Lakatos's methodology of scientific research programs has some dubious features. The single-minded emphasis on progress is questionable. The fact that a series of theories T, T', T'' may be progressing splendidly tells one nothing about whether T'' fits the data well. Why should only the "novel predictions," the new implications of T'' over T', matter? Lakatos's insistence on a specific hard core, which defines a particular research program is also too strict. The supposed "hard core" of every research program is always being reformulated and, in various ways, modified.

If one goes to contemporary philosophy of science in search of hard and fast rules for assessing theories in the light of data, one will be disappointed. Nonphilosophers may find this state of affairs discouraging, and they might draw skeptical or relativist conclusions. But skepticism and relativism are cold comfort when one needs to decide what to do about crushing poverty or the problems of achieving economic growth without environmental disaster. And, as this brief summary shows, philosophers have learned a great deal about theory assessment, even if that knowledge cannot be codified into detailed and exceptionless rules.

The Unity of Science

In studying economics, one not only faces standard problems in the philosophy of science, but one also wants to know whether social sciences like economics should model themselves after natural sciences like physics. Human beings and their social interactions are different objects of study than are planets or proteins. Should the goals and methods of social theorists be the same as those of natural scientists?

As mentioned at the beginning of this introduction, those who have asked whether the social sciences can be "real" sciences have been concerned with several different questions concerning the structure or concepts of theories and explanations in the social sciences and concerning the goals of social theorizing. Philosophers have argued that in addition to or instead of the predictive and explanatory goals of the natural sciences, the social sciences should aim at providing us with *understanding*. This issue receives its classic discussion in the selection in this volume by Max Weber, although Frank Knight also touches on it.[25]

Weber and many others argue that the social sciences should provide understanding "from the inside," that permits social theorists to empathize with the agents and to find what happens "understandable." He argues that social theorists inevitably classify social phenomena in terms of various culturally significant or meaningful categories, and that explanations must be in these terms or they will not tell people what they want to know. This seems to introduce an element of subjectivity into the social sciences that is avoidable in the natural sciences. But, provided that social theorists explain the phenomena in these meaningful terms, Weber has no objection to causal (indeed deductive-nomological) explanation. Yet even here there is a difference in emphasis. Weber maintains that however interested theorists may be in regularities, people want to understand particular happenings in their details and individuality, rather than, as in the natural sciences, as instances of general regularities. I see this as a difference in emphasis, not as demanding a different kind of explanation.

Contemporary philosophers who have been influenced by Weber and by developments in the philosophy of language (especially the work of Wittgenstein), have made stronger claims. These philosophers contend that regularities in human behavior are not natural laws, but the result of *rules* or *institutions.* To "understand" some human action is to discover the rules that guide it. And to understand rules, according to Peter Winch and others, is the same sort of task as understanding meanings. It is a task requiring interpretation, not empirical theorizing and testing. Winch's views seem to rule out applying the methods of the natural sciences to the study of human behavior and institutions, and they have been vigorously contested.[26]

Human free will suggests additional doubts about the possibility of a social science. One wonders whether, given free will, human behavior is intrinsically unpredictable and thus not subject to any laws. As tempting as this line of thought may be, it is a mistake. Even if there are no *deterministic* laws of human behavior, there are, in fact, many regularities in human action. Of course, if Winch and others are right, these regularities differ from laws of nature, but the regularities exist nevertheless. Not only can we predict the behavior of people we know well, but we often know what strangers will do. Every time we cross the street in front of cars stopped for a red light, we stake our lives on such knowledge. Whatever one thinks about free will, there are still uniformities in human behavior, which social theorists may reasonably seek to identify.

The mistaken assertion that human free will makes social science impossible lies, I believe, behind other arguments for the impossibility of any science of society. Expectations and beliefs, including beliefs about social theories, influence behavior. It is thus possible to make both self-fulfilling and self-defeating claims about people. These possibilities suggest that there may be paradoxes lurking within the notion of a social science. But the difficulties are specific and limited rather than fundamental.[27] A social theorist can "factor in" the reactions of those who become aware of any particular theory.

As economists have come increasingly to recognize, human beliefs and expectations, not just the realities about which people have beliefs and expectations are crucial to understanding human behavior. For people can, as Frank Knight points out, make mistakes or fail to recognize things. As a first approximation, economists abstract from such difficulties. They assume that people have perfect information. By assuming that people believe whatever the facts are, economists can avoid worrying about what people actually believe.

Once economists go beyond this first approximation, difficulties arise which have no parallel in the natural sciences. For claims about beliefs (and desires) are, in philosophical jargon, "intentional."[28] They possess a different logic. From a *non*intentional statement such as "The United States invaded

Iraq in 2003," and the second premise, "The invasion of Iraq in 2003 was a huge mistake," one can infer "In 2003, the United States made a huge mistake." But from the same second premise and the *intentional* statement, "President Bush wanted the United States to invade Iraq in 2003," one cannot deduce "President Bush wanted the United States to make a huge mistake." The logic of belief, desire and other such "intentional" terms is in some ways "subjective." These logical peculiarities and the subsequent need for a "subjective" treatment of expectations distinguish economics from the natural sciences (with the possible exception of a small part of biology). However, the significance of the differences is not clear. Members of the Austrian school (represented by James Buchanan and Viktor Vanberg in Chapter 20) argue that these differences are of great importance.

One final special difficulty about the social sciences concerns their role in guiding conduct. One view is that economics serves policy in the same way that the natural sciences guide policies – that is, by helping policy makers to choose means that will achieve their ends. Such a practical role for scientific knowledge seems unproblematic. Agents have some goal that they want to accomplish, and the scientist provides the needed "know-how." On this view, economics matters to policy only as a source of descriptive or "value-free" information. It matters so much, simply because it is so relevant. This view of the policy relevance of economics is defended in many of the essays reprinted in this anthology.

Many disagree. They argue that the links between economics, policy, and values go deeper. The demands and interests of public policy makers or of private employers influence which questions social theorists ask and the range of possible solutions that are seriously considered. The influence can sometimes be crude: economists are people after all, and they can be corrupted by the lure of money and prestige. Or there may be more subtle influences from customs, mores, and rhetoric to avoid what seems "unreasonable" or "irresponsible." Although it is hard to deny that ideological forces have influenced many social scientists, the extent of such ideological and evaluative influences requires sober assessment. What looks like ideology to an unsympathetic critic may in fact be work of unimpeachable intellectual integrity.

There are other less nefarious ways in which economics is entangled in values. Because policy makers rarely turn to economists with precisely formulated goals, economists may help determine the goals. Indeed, as philosophers such as John Dewey have argued, the distinction between means and ends, as plausible and useful as it may sometimes be, may mislead here. The major economists of the past two centuries have also been social philosophers who have found in economic theory inspiration for their social ideals. Although some have argued that normative or welfare economics, which is discussed in Part III of this anthology, is really a part of "positive" economics, investigating means to ends, most would concede that it is driven by moral commitments. Michael S. McPherson and I explore the philosophical foundations of normative economics in Chapter 13.

In providing the reader with both some glimpse of findings in philosophy of science and some sense of how much remains to be found out, this introduction may have discouraged readers who were looking for more detailed guidance. But in recognizing how much there is to be done, readers should not overlook how much has been done. Although logical positivism finds few supporters today, this is because the positivists were so devoted to clarity and precision and so intellectually honest and courageous that they uncovered the inadequacies in their own positions and ultimately refuted themselves. The more historically and empirically oriented philosophy of science and the sometimes exaggerated sociological views that have succeeded them have, no doubt, many inadequacies, but they begin with knowledge that the positivists gained. Similar comments apply to Popper's seminal work.

These words are cold comfort to the citizen, policy maker, economist or social scientist who wants to know whether economics is a science, whether he or she should rely on particular economic theories for practical or theoretical purposes or how he or she can best contribute to economics or to some other social science. But there is nothing to be done other than to make use of what has been learned. Philosophy of science has many insights to offer, and those who do not take it seriously are doomed to repeat its past mistakes. On the basis of such knowledge and on the basis of their own experience, economists and other scientists offer useful rules of thumb. But there is no well-founded general philosophical system to resolve the many real difficulties economists, policy makers, and citizens face.

An Introduction to Economics

To understand the essays collected in this anthology, it helps to know something about economics. What follows does not aim to provide the reader with any technical competence. Its goal is only to give some sense of (a) the basic approach of mainstream economists (b) the different branches of economics and (c) the different schools or approaches of economics.

Although one can find discussions of economics in ancient and medieval philosophy, economics is a modern subject. With the exception of some writing on monetary theory and on the purported benefits of exporting more goods than one imports, economics begins in the eighteenth century with the writings of the French physiocrats, of Cantillon and Hume, and especially of Adam Smith. What set these thinkers apart from the predecessors was their growing recognition of the existence of *mechanisms* whereby individual actions would have systematic consequences without any need for government control of the processes. Smith and others came to see the economy as to a large extent a self-regulating system. Economics came into being when it was realized that there were such things as economic mechanisms and systems to study.

Economics has been concerned mainly with understanding how a capitalist economic system works. (A capitalist economic system is a market economy in which the means of production are for the most part privately owned, and workers are free to accept or decline offers of employment.) Many economists believe that their theories apply to other economic arrangements, too, and a good deal of work has been done on other kinds of economies. But the core of economic theorizing has been devoted to understanding capitalist economies.

Since Adam Smith, a particular vision of such economies has dominated economic theorizing. One conceives of an economy as made up of a large number of independent firms and households, whose interactions with one another consist of voluntary exchanges of goods and services. Everybody knows that people have all sorts of other relations to one another, but the economist assumes as a first approximation that these can be ignored when one is addressing economic problems. Economic agents are conceived of as well-informed, rational, and self-interested agents, with firms seeking to maximize profits and households seeking wealth or what best satisfies their preferences. Agents exchange with one another because they prefer their after-exchange circumstances to their before-exchange circumstances. In the background is an institutional setting that ensures that contracts are kept, violence, coercion and fraud prevented, and so forth. Adam Smith formulates these conditions more loosely than I have, whereas contemporary theorists formulate them much more precisely. But the basic vision has persisted.

Given these assumptions, economists such as Adam Smith have for the most part believed that voluntary exchange would result in an efficient organization of economic life, which would be beneficial to all. In Smith's view, and in the view of most economists since, such a market economy also respects individual liberty more than does any other economic arrangement. One thus has a strong justification for capitalism. It delivers the goods and leaves individuals free to pursue their own objectives. Smith could not, however, prove rigorously that voluntary exchanges of well-informed self-interested agents lead to efficient economic outcomes.

Shortly after World War II, mathematicians and economists such as von Neumann, Arrow, Debreu, and McKenzie proved something like what Smith conjectured. They demonstrated that if agents are rational, self-interested, and well informed, and if they interact only through voluntary exchange in a perfectly competitive market, then a general equilibrium exists, which is Pareto efficient. In addition, they proved that every Pareto efficient outcome is a general equilibrium of voluntary exchanges among rational and self-interested agents, given the proper initial distribution of resources among the agents.[29] A general equilibrium is a situation in which there is no excess demand on any market. An economic outcome *O* is Pareto efficient if and only if one cannot depart from *O* without frustrating someone's preferences. All possibilities for uncontroversial improvement have been seized. In an inefficient economic state of affairs, in contrast, there are ways of better satisfying some people's preferences without lessening the preference satisfaction of others. The "efficiency" in question here is efficiency in satisfying preferences.

Although inefficiency in satisfying preferences is arguably a bad thing, lots of things are worse. Whether a state of affairs is Pareto efficient is generally independent of the distribution of goods, and accordingly some Pareto efficient states of affairs may be intolerable. For example, almost everyone favors a great many nonoptimal economic circumstances over a Pareto efficient state of affairs in which one man had everything he wanted and most others were miserable. One should be skeptical about the significance of proofs of the existence and efficiency of general equilibria both because of the weakness of the notion of Pareto efficiency and because of the extremely restrictive assumptions needed for the proofs.

But I have jumped directly from the beginning to near the end of the story. Let us see how, over the last two centuries, the image of rational, well-informed, and self-interested agents exchanging with one another has been refined. The "classical" economists, of whom Adam Smith, David Ricardo, and John Stuart Mill are the most prominent, did not have much to say about the choices of consumers. Their emphasis was on production and on the factors that influence the supply of consumption goods. They regarded agents as seeking to maximize their financial gains and divided both agents and basic inputs into three major classes: capitalists with their capital (which they conceived of as stocks of accumulated goods or the value thereof), landlords with their land, and workers with their ability to work. The classical economists offered two main generalizations concerning production. First, they assumed that at any given moment all reproducible goods (thus excluding things such as rare paintings) could be produced in any quantity for the same cost per unit. Except for temporary price fluctuations in times of crop failures or rapid changes in demand, prices should be determined by these constant costs of production. Second, classical economists discovered diminishing returns. Unless there is some technological innovation, as more and more labor is devoted to a fixed amount of land, the amount that output increases when an additional laborer is employed will eventually decline.

Given these generalizations concerning production and the view (most forcefully expressed by Malthus) that higher wages cause rapid increases in population, economists in the early nineteenth century drew gloomy conclusions. With economic growth, demand for workers increases and wages rise. The higher wages result in an increase in population. More workers need more food, and so capitalists (whom the classical economists thought of as renting rather than owning land) must rent additional and less fertile land, or they must cultivate existing land more intensively. Either way, the proportional return (rate of profit) on the additional investments will be lower. Landlords will consequently be able to increase rents on more fertile land and the rate of profit throughout the economy must decline. Ricardo argues that eventually the rate of profits will decline to the point where it is no longer worthwhile for capitalists to invest at all. In the resulting "stationary state," there are more workers, but they are no better off than their predecessors, since their wages will decline to that point where population no longer increases. Capitalists are better off than workers, but the rate of profit is low and their returns are modest. The big winners are landlords, who do nothing but collect rents. There is, in the view of most classical economists, little to do about this gloomy prospect except to agitate for the elimination of tariffs impeding the importation of foodstuffs and to preach "restraint" to the working class.

Fortunately, things did not turn out as Ricardo predicted. With improvements in the standard of living, population growth slowed, and by the late nineteenth century, economists recognized that population need not grow explosively in response to higher wages. Moreover, technological improvements brought about increases in productivity beyond the wildest dreams of the classical economists, who vastly underestimated the ability of technological improvements to stave off diminishing returns.

By the end of the nineteenth century, economics was no longer such a dismal science. Economists for the most part stopped worrying about population growth, and through the so-called neoclassical or marginal revolution, they focused their attention on individual choice and exchange. In the 1870s, William Stanley Jevons in England, Carl Menger in Austria, and Leon Walras in France began paying systematic attention to preferences of consumers, to exchange, and to demand for commodities.[30] In doing so, they filled in more of the basic vision of a market economy and transformed economic theory.

Many of the early neoclassical economists, particularly Jevons, were influence by *utilitarianism*, an ethical theory expounded earlier by Jeremy Bentham and John Stuart Mill.[31] According to the utilitarians, questions of social policy are to be answered by calculating the consequences of alternatives for the total happiness of individuals. The policy that maximizes the sum of individual utilities is the morally right one. Bentham held that the utility of something to an individual is a sensation that might in principle be quantified and measured. He also believed that individuals act so as to maximize their own utility (which raises the question of how they can be motivated to carry out actions that instead maximize the sum of everybody's utility).

Jevons developed the essentially Benthamite notion of a utility function. Every option open to an individual results in a certain amount of utility for that person. One can then clarify the notion of rationality by maintaining that people act so as to maximize some consistent utility function. In addition, the neoclassical economists assumed that consumers are generally not satiated – that they will always prefer a bundle x of commodities or services to another bundle y if x is unambiguously larger than y. Nonsatiation is both a plausible first approximation, and it articulates the notion of self-interest. All that matters to agents are the bundles of commodities and services that they are giving up or receiving.

With the addition of one more generalization, one has the core of modern economic theory. The early neoclassical economists noted that as one consumes more of any commodity or service, each additional unit increases one's utility at a diminishing rate. One's first computer may raise one's utility considerably. A second computer doesn't contribute nearly as much. This law of diminishing marginal utility explains why the price of essential but plentiful commodities such as water is lower than the price of inessential but scare commodities such as diamonds. Thinking in terms of marginal utility also enables one to give an integrated account of the "forces" affecting both demand and supply. Instead of regarding costs as reflecting physical requirements, most neoclassical theorists take costs to be the disutilities incurred when individuals devote resources or service to production or to be the utilities that would result from alternative uses of resources that individuals forgo (although these are in turn influenced by technical factors). The forces governing supply and demand are ultimately the same. The role of the market is to equilibrate these forces and to bring into harmony the efforts of individuals to secure what they want. With the further simplification that commodities are infinitely divisible, it became possible to apply the calculus to economics and to formulate this theory mathematically. In principle, the single theory of general equilibrium should enable one to explain virtually all the significant features of an economy.

In the 130 years since the neoclassical revolution, this theory has been tremendously refined. In speaking of utility, for example, contemporary economists are no longer speaking of some sensation that individuals want to maximize. "Utility" is now just another way to speak about preferences. The utility of some object of choice *x* to agent *A* is larger than that of option *y* if and only if *A* prefers *x* to *y*. In taking utility to reflect merely the ordering of preferences, economists had to surrender talk of utility differences and hence of marginal utility. Fortunately, the law of diminishing marginal utility can be reformulated in terms of the diminishing rates with which individuals are willing to substitute units of one commodity for another. Roughly speaking, one can replace the "law" of diminishing marginal utility with the generalization that people are willing to pay less for additional units of commodities that they already have a lot of than for commodities that they have very little of. Despite these refinements, mainstream theory is still recognizably the theory developed by the early neoclassical economists.

This fact may seem surprising, as most people know that in response to the Great Depression of the 1930s, John Maynard Keynes proposed a dramatic overhaul of economics. Before Keynes, most theorists of any reputation had maintained that a prolonged depression was impossible. There might be a crisis of confidence, which would lead to a temporary hoarding of money and a temporary interruption in the general cycle of exchange (in which firms as a whole purchase resources from their owners, then sell the commodities produced to the latter in their role as consumers, who then sell resources to firms again and so on). But with an excess demand for money and excess supplies of resources and commodities, prices are bound to drop and real interest rates rise. Any tendency to hoard would be self-correcting.

Keynes challenged this orthodoxy in part by emphasizing the importance of liquidity to both firms and individuals when they are faced with the uncertainties that a business crisis causes, and in part because he questioned the efficacy of the supposed self-correcting mechanisms. Prices, especially wages, do not drop easily, and lower wages can lead to less spending, which would suppress demand for commodities and lead to an even deeper slump. Keynes argued that government policy could increase aggregate demand for commodities and encourage investment and in that way move the economy out of its unemployment "equilibrium."

Despite Keynes's influence, his work did not shake the fundamentals of neoclassical theory. Initially, neoclassical theory instead divided into microeconomics, on the one hand, which is concerned with individuals, firms, and industries, and macroeconomics, on the other hand, which is concerned with aggregate demand and the performance of the economy as a whole. Although vestiges of this bifurcation persist, there ought, one would think, to be important connections between microeconomics and macroeconomics, and most economists nowadays insist on relating macroeconomic theories to stylized microeconomic foundations. For further discussion, see Chapter 17.

In the 1970s and 1980s, Keynesian economics was seriously challenged. Not only was it ill-suited to deal with the simultaneous inflation and unemployment of the 1970s, but economists grew increasingly impatient with the gap between micro- and macroeconomics and increasingly enamored of microeconomics. Unlike previous economists who made use of Keynesian macroeconomics while hoping to reconcile it with microeconomics, members of the so-called new classical school refused to employ any models that did not at least purport to derive from microeconomics or general equilibrium theory. Some, such as Robert Lucas, even went so far as to deny on the basis of microeconomic considerations that there was such a thing as involuntary unemployment,[32] and Lucas and others argued that the rational expectations of economic agents tend to undermine the effectiveness of monetary and fiscal policy as tools to manage the economy.

It is hard to say whether the new classical research program has triumphed or failed. On the one hand, its econometric predictions were no improvement over its predecessors, and the experience of the 1990s made it hard to believe that policy (especially monetary policy) had only a very limited effect on the economy. Updated versions of Keynesian economics remain influential. On the other hand, the concerns about modeling rational expectations that the new classical economists emphasized are now widely accepted, and new classical economics lives on in a different form as so-called real business cycle theory, which argues that business cycles are largely a response to "real" as opposed to monetary or policy factors.[33] Variations on real business cycle models are currently very influential. As this brief description suggests, macroeconomics is an unsettled area of economics.

Although microeconomics and macroeconomics are the two main branches of mainstream economics, they do not include all of it. Over the past three generations, there has been an enormous expansion of *econometrics*, which is discussed in Chapter 16. Econometrics is a branch of applied statistics as well as a branch of economics. Beginning in the 1930s, it was hoped that the claims of economic theorists might be tested and refined with the help of statistical techniques. Since then econometric techniques have become much more sophisticated. Exactly what this work means for economic theory (as opposed to narrowly focused practical inquiries) is controversial, with some prominent economists arguing that econometrics is incapable of providing good reasons to believe or disbelieve any significant causal claims.[34]

Microeconomics, macroeconomics, and econometrics together include most of mainstream economics, although there are of course specific sub-areas such as international trade, labor economics, and so forth. There are also competing schools of economics, although in most cases they have relatively few proponents. A generation ago, there was still a good deal of interest in Marxian economics. Although Marx was heavily influenced by Ricardo's work, he had a different view of the nature of economics and of its relationship to other social sciences than classical or neoclassical economists have. According to Marx's *historical materialism* (which is sketched in Chapter 5), the relations among people in the course of their productive activities are the most fundamental social relations. Relations of production strongly influence not only other relationships but also the personalities and consciousness of individuals. In studying economics, one is studying much more than how individuals produce, exchange, and distribute goods and services; one is also studying how human beings shape the development of their species.

Marx regards capitalism, despite the miseries it may cause (which he meticulously documents), as an enormous step forward for human beings. Capitalism relates individuals everywhere to one another through the world market, and it expands the needs and horizons of people. But, as argued in his early essay "Estranged Labor," it does not allow people to decide rationally and consciously how society and human nature should develop. The market creates both a reality and an illusion of helplessness. Given the market, people cannot in fact consciously determine their collective future. At the same time, it is an illusion to regard capitalism as eternal or natural. Marx believes that people can and will transcend capitalism and organize production and distribution in some rational way.

Those who will carry out the socialist revolution are the workers, who are "exploited," because they do all the producing but receive only part of the output. Capitalists (who, on Marx's view, possess little more real freedom than do workers) would resist any revolution that attempts to take their property and to prevent them from hiring workers and making profits. But, or so Marx argues, in expanding the size of their enterprises, capitalists unwittingly enlarge and strengthen the working class and lay the foundations for socialist revolution.

Given the collapse of the Soviet Union and the transformation of its economy and of the economies of Eastern Europe, interest in Marxian economics has collapsed as well. From one perspective, this is peculiar, because the Russian and Eastern European economies had only a tenuous connection with Marx's economics. But economic theories do not hover above the political waves. They are instead tossed about and, in the case of Marx's economics, possibly drowned.

Institutionalist (or "evolutionary") economists make up another major contemporary alternative to mainstream economics. The neoclassical attempt to capture all relevant aspects of the economy in one elegant theory leaves out a great deal. In the view of the institutionalists (and their nineteenth-century predecessors, the German Historical School), it leaves out too much and in abstracting from institutional development, it misses central aspects. The essays by Thorstein Veblen (Chapter 6) and Geoffrey M. Hodgson (Chapter 21) exemplify the institutionalist critique of mainstream economics and provide some sense of the institutionalist alternative. Although institutionalists do not ignore individual decision making, they emphasize the evolving constraints on agents occupying specific economic roles. The institutionalists do not constitute a tightly organized sect. The writings of the central historical figures (Thorstein Veblen, Wesley Mitchell, and John R. Commons) are very different from one another. The emphasis is on historically situated and evolutionary theorizing. Economists are divided on how successful institutional theorizing has been and is likely to be.

A third contemporary alternative to mainstream economics, about which there is currently heated disagreement, is behavioral economics, including neuroeconomics.[35] Behavioral economics has been heavily influenced by the increasingly important experimental work that economists have been doing, which is discussed in Chapter 18 by Vernon Smith. The general dissatisfaction many economists have felt with the highly simplified assumptions that mainstream economics makes concerning individual beliefs and preferences has been superseded by carefully delineated behavioral anomalies that have been established through economic experimentation. It is now possible to study the influence on preferences of a wide variety of cognitive, motivational, and even neurological features of human beings and to develop theories of economic behavior that are more psychologically nuanced. Whether and to what extent this work will help economists to address the questions concerning monetary policy, tax incidence, or economic welfare are hotly contested matters.[36]

These are but three of many approaches, which, in addition to mainstream neoclassical economics, occupy contemporary economists. A few others deserve to be mentioned. Neo-Ricardians believe that one can do better in understanding economies by employing modern mathematical reformulations and extensions of Ricardo's economics than by employing its neoclassical successor.[37] Austrian economists agree with neoclassical economists on the central generalizations of economics, but stress the importance of uncertainty, disequilibrium, and a subjective point of view (Chapter 20). Because of these factors, they regard sophisticated mathematical analyses of equilibria as misleading.[38] Post-Keynesian economists often offer similar criticisms of high theory, but unlike the Austrians, they tend to defend interventionist policies.[39] Economic forecasters often depend very little on any specific economic theory. And the list could be extended. Although contemporary economics is dominated by mainstream microeconomics, macroeconomics, and econometrics, there is lots more going on.

An Introduction to Economic Methodology

John Stuart Mill's 1836 essay, with which this analogy begins, is one of the first discussions of the methodology of economics, and it is still one of the best. From the perspective of a staunch empiricist like Mill, economics is a puzzling science. Its conclusions, which Mill accepts, are rarely tested, and they sometimes appear to be disconfirmed. Specific predictions based on economic theory are inexact and sometimes dead wrong. How can Mill reconcile his confidence in economics and his empiricism?

In Mill's view, the basic premises of economics are either psychological claims, which are established by introspection, or technical claims, such as the law of diminishing returns, which are established directly by experimentation. These premises state how specific causal factors operate. If the only causal factors that affect economics were those that economists consider, then the conclusions of economics would be correct, because they follow deductively from its well-supported premises. In fact, Mill argues, the conclusions economists draw must be treated cautiously, because so much is left out of their theory. Economists must be ready to make allowances for various disturbances, and economists must recognize that their predictions may be badly mistaken even though their theory is fundamentally correct. They should regard economics as hypothetical – as a science of tendencies, whose influence may be overwhelmed by interferences. Because it is only a science of tendencies, economists and policy makers cannot be confident that its predictions are always correct.

Mill's view was influential throughout the nineteenth and early twentieth century. It is, for example, still alive in John Neville Keynes's authoritative summing up in *The Scope and Method of Political Economy*, excerpts from which were reprinted in the first edition of this anthology. Despite differences in language, tone, and emphasis, Weber adopts a similar position in his discussion of "ideal types" in Chapter 2.

The transition from classical to neoclassical economics brought substantial changes in economic doctrine and changes in methodology. In its focus on individual decision making, neoclassical theory is a more individualist and subjective theory than was its classical predecessor, and the appreciation of this fact is an important contribution of early twentieth-century methodological writing. The major figures in developing this subjective turn are the Austrians (including especially von Mises), Frank Knight and Lionel Robbins. Knight's distinctive methodological contribution lies in his distinction between risk, on the one hand, (where the alternatives and their probabilities are known) and error and true uncertainty, on the other hand. Knight and the Austrians agree that as soon as one abandons the subjective point of view and thinks of economics as if it were a natural science, one loses sight of the central features of the subject.

Lionel Robbins, in his classic *An Essay on the Nature and Significance of Economic Science* (Chapter 3), comes close to the view of the Austrians, but he is better known for his definition of economics as "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses" (1935, p. 85). According to this definition, economics is not concerned with any particular class of social phenomena (such as production, distribution, exchange, or consumption). Economics is instead concerned with a particular aspect of human behavior. One's decisions to have children or to be unfaithful to one's spouse are, on this definition, as much a part of economics as supply and demand for tuna. Robbins is, in effect, defining economics as the science of rational choice – that is, as neoclassical theory. Such redefinitions are characteristic of scientific development.[40] Robbins's definition remains controversial, since it excludes from economics some work that most people regard as economics, such as Keynesian theory.

Robbins, Knight, and the Austrians stress the individualism and subjectivity of economics, and they all emphasize the peculiarities of human action as an object of scientific investigation. They also agree with Mill that the basic premises of economics are well established and that these premises are not impugned by the empirical failures of the theory. In fact, the Austrians go further and argue that the basic premises are *a priori* truths.

With the intrusion of the views of the logical positivists in the 1930s came the first important change in the profession's views on the justification of economic theory. In 1938, Terence Hutchison published *The Significance and Basic Postulates of Economic Theory*. In this landmark book, Hutchison argues that economics, like other sciences, must formulate testable generalizations and subject them to serious tests. The statements of "pure theory" in economics are, Hutchison argues, empty definitional truths. Claims in economics are so hedged with *ceteris paribus* qualifications that they are untestable. With the weight of contemporary logical positivism behind him, Hutchison insisted that it was time for economists to start behaving like responsible scientists. The development of revealed-preference theory and Paul Samuelson's defense of what he calls "operationalism" also supported the demand that economics be recast into testable theories.

Hutchison's criticisms were immediately rebutted by economists such as Knight, but they remained disturbing. Could it be that economics did not meet the standards of empirical science? Some, such as Knight and the Austrians, were prepared to say that the standards of the natural sciences did not apply to economics. But most writers on economic methodology attempted to show that economics satisfied the more sophisticated (and weaker) criteria to which the logical positivists had already retreated.[41] Although Milton Friedman's well-known essay, "The Methodology of Positive Economics" (1953; see Chapter 7 in this anthology) does not refer to contemporary philosophy of science, it, too, attempts to show that economics satisfies broadly positivist standards.

For decades after its publication, Friedman's essay dominated work on the methodology of economics. Although almost all the many essays that have been written in response to it have been critical (like the brief comments in Chapters 8 and 9 of this volume), Friedman's essay has nevertheless remained the most influential work on economic methodology of the twentieth century.

One should not forget that there are many different methodological questions that one can ask about economics. The different branches and schools of economics face special methodological problems of their own, which are discussed in the six essays reprinted in Part IV of this anthology. Questions concerning the relations between positive and normative and the character of normative economics are the topic of Part III, although selections in other sections bear on this issue, too.

The field of economic methodology, including methodological studies of the details of branches and schools of economics has blossomed during the last fifteen years. Part V turns to some of the new directions within economic methodology, and the widespread changes in the contents of the other parts of this anthology reflect this blossoming. The extent to which the field has matured was brought home to me vividly by how hard it was to decide on what to include. In the first edition of this anthology, I noted that at least nineteen books specifically devoted to economic methodology had been published in English between 1975 and 1983. In the decade between the first and the second edition, I counted fifty. Since the second edition, there have been about one hundred more, and the outpouring of essays has increased at a greater pace. Just after the first edition of this anthology was published, a new journal, *Economics and Philosophy* began publishing works on methodology, the theory of rationality, and ethics and economics. Just before the second edition of this anthology came out, the *Journal of Economic Methodology* began publishing essays and reviews specifically focused on methodology. And the pace of publication of essays on economic methodology in journals in economic theory, philosophy of science, and history of economics or history of science has increased rapidly, too. Were it not for the generous advice of many others, who are expert in particular sub-domains of economic methodology, I would not have been able to do a competent job of designing this edition of the anthology.[42] The literature is now just too large!

The methodological questions economics raises are varied, difficult, and for the most part unanswered. When I compiled the first edition of this anthology, I was optimistic that collaboration between philosophers and economists would tame, if not answer, these questions. To some extent that optimism has been rewarded: progress has been enormous. Just compare the essays in a current version of *Economics and Philosophy* or *The Journal of Economic Methodology* with the essays in the early issues of either journal. I would like to think that this anthology, now in its third edition and its third decade, has contributed to that progress.

I am perhaps a little less optimistic now (or perhaps just older). The methodological problems economics raises are difficult, and progress is slow when philosophical argument has to contend with social forces and the reward structure within academic disciplines. There is so much more to be learned about the nature of economic models, how to compare and assess them, how to relate them to policy recommendations and empirical studies, and, most important, how to improve them. May this third edition continue to play a role in tackling these questions.

Notes

1. For defenses of scientific realism, see R. Boyd, "On the Current Status of Scientific Realism," *Erkenntnis* 19 (1983): 45–90 and R. Miller, *Fact and Method: Explanation, Confirmation and Reality in the Natural and the Social Sciences* (Princeton: Princeton University Press, 1987), Part III. For an influential antirealist position, see B. van Fraassen, *The Scientific Image* (Oxford: Oxford University Press, 1980).
2. See C. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (New York: Free Press, 1965), and for an overview of debates about scientific explanation, W. Salmon, *Four Decades of Scientific Explanation* (Minneapolis: University of Minnesota Press, 1990).
3. This is Wesley Salmon's example from "Statistical Explanation," in W. Salmon, ed. *Statistical Explanation and Statistical Relevance* (Pittsburgh: University of Pittsburgh Press, 1971), p. 34.
4. The example is adapted from S. Bromberger, "Why Questions," in R. Colodny, ed. *Mind and Cosmos: Essays in Contemporary Science and Philosophy* (Pittsburgh: University of Pittsburgh Press, 1966), pp. 86–111.
5. In both *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983) and *Nature's Capacities and Their Measurement* (Oxford: Clarendon Press, 1989), Nancy Cartwright argues compellingly for the importance of causal notions in science, particularly physics. Explicitly causal accounts of explanation are developed in D. Lewis, "Causal Explanation," pp. 214–40 of his *Philosophical Papers*, vol. 2 (Oxford: Oxford University Press, 1986), R. Miller, *Fact and Method*, Part 2, and W. Salmon, *Scientific Explanation and the Causal Structure of the World*. The most sophisticated current account of causal explanation is in J. Woodward, *Making Things Happen* (Oxford: Oxford University Press, 2003).
6. See G. von Wright, *Explanation and Understanding* (Ithaca, NY: Cornell University Press, 1971) and, for a more introductory treatment, chapter 2 of A. Rosenberg, *Philosophy of Social Science* (Boulder, CO: Westview Press, 1988).
7. See D. Davidson, "Actions, Reasons and Causes," *Journal of Philosophy* 60 (1963): 685–700 and A. Rosenberg, *Microeconomic Laws: A Philosophical Analysis* (Pittsburgh: University of Pittsburgh Press, 1976), ch. 5.
8. Although the logical positivists and logical empiricists also talked about models, their notion, which was borrowed from formal logic, has little to do with the models that economists employ.
9. See F. Suppe, *The Structure of Scientific Theories*, 2nd ed. (Urbana: University of Illinois Press, 1977), esp. pp. 3–118.
10. This predicate view of models derives from P. Suppes, *Introduction to Logic* (New York: Van Nostrand Reinhold, 1957), ch. 12, J. Sneed, *The Logical Structure of Mathematical Physics* (Dordrecht: Reidel, 1971), and W. Stegmueller, *The Structuralist View of Theories* (New York: Springer-Verlag, 1979). For a convenient introductory treatment that follows the terminology that I prefer, see R. Giere, *Understanding Scientific Reasoning*, 2nd ed. (New York: Holt, Rinehart and Winston, 1982), ch. 5. These authors all develop this view as an account of theories rather than models. For a more detailed exposition of the view with reference

to economics, see my *The Inexact and Separate Science of Economics* (Cambridge: Cambridge University Press, 1992), ch. 5.

11. See Ludwig von Mises, *Human Action: A Treatise on Economics* (New Haven: Yale University Press, 1949). Other aspects of the Austrian approach are developed in Chapter 20.
12. The classic critique of the analytic-synthetic distinction is W. V. O. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View*, 2nd ed. (New York: Harper & Row, 1961), pp. 20–46. See also H. Putnam, "The Analytic and the Synthetic," in H. Feigl and G. Maxwell, *Minnesota Studies in the Philosophy of Science*, vol. 3, pp. 350–97, and C. Hempel, "A Logical Appraisal of Operationalism," in his *Aspects of Scientific Explanation*, pp. 123–33.
13. *An Inquiry Concerning Human Understanding* (1748; rpt. Indianapolis: Bobbs-Merrill, 1955), p. 47.
14. In rejecting foundationalism, I am following many contemporary philosophers. Recent influential antifoundationalists include Quine in "Two Dogmas," and "Epistemology Naturalized," in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), pp. 69–90 and I. Levi in *The Enterprise of Knowledge* (Cambridge, MA: MIT Press, 1980). The antifoundationalism of Quine and Levi derives in large part from the American pragmatists.
15. R. Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).
16. For an excellent overview of recent work on confirmation, see the entry on "Confirmation Theory" by P. Maher in *The Encyclopedia of Philosophy*, 2nd ed., edited by D. Borchert (London: Macmillan, 2005). For an overview of issues concerning theory appraisal with special reference to economics, see the entry by E. Eells and D. Hausman in the second edition of the *New Palgrave Dictionary of Economics* (London: Macmillan, 2007).
17. See K. Popper, *The Logic of Scientific Discovery* [1935] (London: Hutchinson & Co., 1959).
18. Contrast K. Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford: Clarendon Press, 1972), ch. 1 with *The Logic of Discovery*, ch. 2, esp. p. 50. Despite this disavowal, Popper continues to stress the asymmetry between falsification and verification.
19. K. Popper, "Science: Conjectures and Refutations," in *Conjectures and Refutations: The Growth of Scientific Knowledge*, 3rd ed. (London: Routledge, 1969), pp. 33–65.
20. *The Logic of Scientific Discovery*, ch. 5. For contrasting views of the merits of Popper's philosophy of science and their applicability to economics, see B. Caldwell, "Clarifying Popper." *Journal of Economic Literature* 29 (1991): 1–33, D. W. Hands, *Testing, Rationality and Progress* (Lanham, MD: Rowman and Littlefield, 1992), ch. 11, and my *The Inexact and Separate Science of Economics*, ch. 10.
21. T. Kuhn, *The Structure of Scientific Revolutions* 2nd ed. (Chicago: University of Chicago Press, 1970), I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," rpt. in his *Philosophical Papers* (Cambridge: Cambridge University Press, 1978), vol. 1, pp. 8–101.
22. Especially in chapters 9 and 10 of *The Structure of Scientific Revolutions.*

23. See J. Brown, ed., *Scientific Rationality: The Sociological Turn* (Dordrecht: Kluwer, 1984) and, for an overview with applications to economics, D. W. Hands, *Reflection without Rules: Economic Methodology and Contemporary Science Theory* (Cambridge: Cambridge University Press, 2001), chapter 5 and U. Mäki, "Social Conditioning of Economics," in Neil deMarchi, ed. *Post-Popperian Methodology of Economics: Recovering Practice* (Boston: Kluwer, 1994), pp. 65–104.
24. See "Falsification and the Methodology of Scientific Research Programmes," pp. 116–38, and especially pp. 124n, 125n, and 175n.
25. For a modern collection of views concerning "understanding" in the social sciences, see F. Dallmayr and T. McCarthy, eds., *Understanding and Social Inquiry* (Notre Dame: University of Notre Dame Press, 1977).
26. See P. Winch, *The Idea of a Social Science and its Relation to Philosophy* (London: Routledge, 1958). Other prominent figures arguing that explanations in the social sciences involve interpretation are G. Anscombe, P. Geach, A. Meldon, and G. von Wright.
27. These general philosophical issues are linked to controversies concerning rational expectations in economics. See K. Hoover, *The New Classical Macroeconomics: A Sceptical Inquiry* (Oxford: Basil Blackwell, 1988).
28. See, for example, R. Chisholm, *Perceiving* (Ithaca, NY: Cornell University Press, 1957), ch. 11 and W. Quine, *Word and Object* (Cambridge, MA: MIT Press, 1960), chs. 4, 6.
29. For the bright spots in this history, see E. Roy Weintraub, *General Equilibrium Analysis: Studies in Appraisal* (Cambridge: Cambridge University Press, 1985). For some of the less successful aspects, see Weintraub's *Stabilizing Dynamics* (Cambridge: Cambridge University Press, 1991).
30. W. Jevons, *The Theory of Political Economy* (1871); C. Menger, *Grundsätze der Volkwirthschaftslehre* (1871); and L. Walras *Elements of Pure Economics* (1874), tr. W. Jaffé (1954). Many others were important, too, especially Alfred Marshall, whose *Principles of Economics* (1st ed., 1890) was for decades the main text of neoclassical economics. The common features in the work of Jevons, Menger, Walras, and Marshall or in the work of their immediate intellectual descendants should not be exaggerated. There also were sharp differences.
31. See particularly Mill's *Utilitarianism* (1863).
32. R. Lucas, "Unemployment Policy," *American Economic Review* 68 (1978): 353–57.
33. For an excellent overview, see G. Stadler, "Real Business Cycles," *Journal of Economic Literature*, 32 (1994): 1750–83.
34. L. Summers, "The Scientific Illusion in Empirical Macroeconomics," *Scandinavian Journal of Economics* 93 (1991): 129–48.
35. For an overview, see C. Camerer, G. Loewenstein, and M. Rabin, eds. *Advances in Behavioral Economics* (Princeton: Princeton University Press, 2004).
36. See particularly, A. Caplin and A. Schotter, eds., *Perspectives on the Future of Economics: Positive and Normative Foundations* (Oxford: Oxford University Press, 2007).
37. Leading neo-Ricardian theorists include J. Eatwell, P. Garegnani, and L. Pasinetti. The major theoretical work influencing the neo-Ricardians is P. Sraffa, *The*

Production of Commodities by Means of Commodities (Cambridge: Cambridge University Press, 1960).

38. Prominent Austrian economists include L. von Mises, F. von Hayek, and M. Rothbard.
39. See Alfred Eichner, ed., *A Guide to Post-Keynesian Economics* (White Plains, NY: M.E. Sharpe, 1978), and the *Journal of Post-Keynesian Economics.*
40. See Kuhn, *Structure of Scientific Revolutions*, chapters 9 and 10.
41. Fritz Machlup's work, some of which appeared in the first two editions of this anthology is particularly noteworthy in this regard. See F. Machlup, *Methodology of Economics and Other Social Sciences* (New York, Academic Press, 1978). I have argued that Mill's views are useful in allaying such empiricist qualms. See *The Inexact and Separate Science of Economics* (Cambridge: Cambridge University Press, 1992), esp. chs. 8 and 12.
42. I would in particular like to thank Roger Backhouse, Peter Boettke, Richard Bradley, Bruce Caldwell John Davis, Francesco Guala, Wade Hands, Kevin Hoover, Harold Kincaid, Uskali Mäki, Deirdre McCloskey, Michael McPherson, Philippe Mongin, Julie Nelson, Alex Rosenberg, Don Ross, and Margaret Schabas, to whom I sent appeals for help. Wade Hands, Kevin Hoover, and Elliott Sober read drafts of this introduction and helped improve it. One of great privileges of having worked so long on economic methodology is being able to count such wonderful people and wonderful intellects as friends.

PART ONE

CLASSIC DISCUSSIONS

The six selections reprinted in this section are a good sample of the major contributions to the philosophy and methodology of economics before the late 1930s, when logical positivism became influential. Not all the significant works could be included – even in abridged form – but many of the methodological insights of authors omitted here, such as J. E. Cairnes, J. N. Keynes, Carl Menger, W. S. Jevons, Alfred Marshall, and Ludwig von Mises appear in other essays in this anthology.

The materials collected in this section represent a number of different perspectives and have stood the test of time. Although economic theory has changed considerably since Mill or Marx or Veblen wrote, their appreciation of the methodological difficulties of economics still rewards careful study. One might, in fact, argue that thinking on economic methodology has advanced very little beyond the stage to which the authors in this section brought it.

ONE

On the Definition and Method of Political Economy

John Stuart Mill

John Stuart Mill (1806–73) was born in London, His father, James Mill, was a friend of Bentham and of Ricardo and did important work himself in psychology and political science. As John Stuart Mill explains in his autobiography, he was educated at home by his father, starting Greek at age 3 and Latin at age 8. By age 13 Mill had been through a complete course in political economy. Mill spent most of his life working for the East India Company. His *Principles of Political Economy* (1848) was the nineteenth century's most influential text in economics, and his *A System of Logic* (1843) was the century's most influential text in logic and the theory of knowledge. His essays on ethics and contemporary culture, such as *Utilitarianism* and *On Liberty*, continue to be extremely influential. Mill was an early defender of women's rights and of a moderate democratic socialism. The following selection is an abridgment of Mill's "On the Definition of Political Economy and the Method of Investigation Proper to It." Approximately the first quarter of the essay, in which Mill discusses the definition of economics, is omitted.

What is now commonly understood by the term "Political Economy" is not the science of speculative politics, but a branch of that science. It does not treat of the whole of man's nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, aversion to labour, and desire of the present enjoyment of costly indulgences. These it takes, to a certain

---

Excerpted from "On the Definition of Political Economy and the Method of Investigation Proper to It" (1836). Reprinted in *Essays on Some Unsettled Questions of Political Economy* (1844), 3d ed., London: Longmans Green & Co., 1877, pp. 120–64.

extent, into its calculations, because these do not merely, like other desires, occasionally conflict with the pursuit of wealth, but accompany it always as a drag, or impediment, and are therefore inseparably mixed up in the consideration of it. Political Economy considers mankind as occupied solely in acquiring and consuming wealth; and aims at showing what is the course of action into which mankind, living in a state of society, would be impelled, if that motive, except in the degree in which it is checked by the two perpetual counter-motives above adverted to, were absolute ruler of all their actions. Under the influence of this desire, it shows mankind accumulating wealth, and employing that wealth in the production of other wealth; sanctioning by mutual agreement the institution of property; establishing laws to prevent individuals from encroaching upon the property of others by force or fraud; adopting various contrivances for increasing the productiveness of their labour; settling the division of the produce by agreement, under the influence of competition (competition itself being governed by certain laws, which laws are therefore the ultimate regulators of the division of the produce); and employing certain expedients (as money, credit, &c.) to facilitate the distribution. All these operations, though many of them are really the result of a plurality of motives, are considered by Political Economy as flowing solely from the desire of wealth. The science then proceeds to investigate the laws which govern these several operations, under the supposition that man is a being who is determined, by the necessity of his nature, to prefer a greater portion of wealth to a smaller in all cases, without any other exception than that constituted by the two counter-motives already specified. Not that any political economist was ever so absurd as to suppose that mankind are really thus constituted, but because this is the mode in which science must necessarily proceed. When an effect depends upon a concurrence of causes, those causes must be studied one at a time, and their laws separately investigated, if we wish, through the causes, to obtain the power of either predicting or controlling the effect; since the law of the effect is compounded of the laws of all the causes which determine it. The law of the centripetal and that of the tangential force must have been known before the motions of the earth and planets could be explained, or many of them predicted. The same is the case with the conduct of man in society. In order to judge how he will act under the variety of desires and aversions which are concurrently operating upon him, we must know how he would act under the exclusive influence of each one in particular. There is, perhaps, no action of a man's life in which he is neither under the immediate nor under the remote influence of any impulse but the mere desire of wealth. With respect to those parts of human conduct of which wealth is not even the principal object, to these

Political Economy does not pretend that its conclusions are applicable. But there are also certain departments of human affairs, in which the acquisition of wealth is the main and acknowledged end. It is only of these that Political Economy takes notice. The manner in which it necessarily proceeds is that of treating the main and acknowledged end as if it were the sole end; which, of all hypotheses equally simple, is the nearest to the truth. The political economist inquires, what are the actions which would be produced by this desire, if, within the departments in question, it were unimpeded by any other. In this way a nearer approximation is obtained than would otherwise be practicable, to the real order of human affairs in those departments. This approximation is then to be corrected by making proper allowance for the effects of any impulses of a different description, which can be shown to interfere with the result in any particular case. Only in a few of the most striking cases (such as the important one of the principle of population) are these corrections interpolated into the expositions of Political Economy itself; the strictness of purely scientific arrangement being thereby somewhat departed from, for the sake of practical utility. So far as it is known, or may be presumed, that the conduct of mankind in the pursuit of wealth is under the collateral influence of any other of the properties of our nature than the desire of obtaining the greatest quantity of wealth with the least labour and self-denial, the conclusions of Political Economy will so far fail of being applicable to the explanation or prediction of real events, until they are modified by a correct allowance for the degree of influence exercised by the other cause.

Political Economy, then, may be defined as follows: and the definition seems to be complete:

> The science which traces the laws of such of the phenomena of society as arise from the combined operations of mankind for the production of wealth, in so far as those phenomena are not modified by the pursuit of any other object.

But while this is a correct definition of Political Economy as a portion of the field of science, the didactic writer on the subject will naturally combine in his exposition, with the truths of the pure science, as many of the practical modifications as will, in his estimation, be most conducive to the usefulness of his work.

The above attempt to frame a stricter definition of the science than what are commonly received as such, may be thought to be of little use; or, at best, to be chiefly useful in a general survey and classification of the sciences, rather than as conducing to the more successful pursuit of the particular science in question. We think otherwise, and for this reason; that, with the consideration of the definition of a science, is inseparably connected *that of the philosophic method* of the science; the nature of the process by which its investigations are to be carried on, its truths to be arrived at.

Now, in whatever science there are systematic differences of opinion – which is as much to say, in all the moral or mental sciences, and in Political Economy among the rest; in whatever science there exist, among those who have attended to the subject, what are commonly called differences of principle, as distinguished from differences of matter-of-fact or detail, – the cause will be found to be, a difference in their conceptions of the philosophic method of the science. The parties who differ are guided, either knowingly or unconsciously, by different views concerning the nature of the evidence appropriate to the subject. They differ not solely in what they believe themselves to see, but in the quarter whence they obtained the light by which they think they see it.

The most universal of the forms in which this difference of method is accustomed to present itself, is the ancient feud between what is called theory, and what is called practice or experience. There are, on social and political questions, two kinds of reasoners: there is one portion who term themselves practical men, and call the others theorists; a title which the latter do not reject, though they by no means recognize it as peculiar to them. The distinction between the two is a very broad one, though it is one of which the language employed is a most incorrect exponent. It has been again and again demonstrated, that those who are accused of despising facts and disregarding experience build and profess to build wholly upon facts and experience; while those who disavow theory cannot make one step without theorizing. But, although both classes of inquirers do nothing but theorize, and both of them consult no other guide than experience, there is this difference between them, and a most important difference it is: that those who are called practical men require *specific* experience, and argue wholly *upwards* from particular facts to a general conclusion; while those who are called theorists aim at embracing a wider field of experience, and, having argued upwards from particular facts to a general principle including a much wider range than that of the question under discussion, then argue *downwards* from that general principle to a variety of specific conclusions.

Suppose, for example, that the question were, whether absolute kings were likely to employ the powers of government for the welfare or for the oppression of their subjects. The practicals would endeavour to determine this question by a direct induction from the conduct of particular despotic monarchs, as testified by history. The theorists would refer the question to be decided by the test not solely of our experience of kings, but of our experience of men. They would contend that an observation of the tendencies which human nature manifested in the variety of situations in which human beings have been placed, and especially observation of what passes in our own minds, warrants us in inferring that a human being in the situation of a despotic king will make a bad use of power; and this conclusion would lose nothing of its certainty even if absolute kings had never existed, or if history furnished us with no information of the manner in which they had conducted themselves.

The first of these methods is a method of induction, merely; the last a mixed method of induction and ratiocination. The first may be called the method *à posteriori*; the latter, the method *à priori*. We are aware that this last expression is sometimes used to characterize a supposed mode of philosophizing, which does not profess to be founded upon experience at all. But we are not acquainted with any mode of philosophizing, on political subjects at least, to which such a description is fairly applicable. By the method *à posteriori* we mean that which requires, as the basis of its conclusions, not experience merely, but specific experience. By the method *à priori* we mean (what has commonly been meant) reasoning from an assumed hypothesis; which is not a practice confined to mathematics, but is of the essence of all science which admits of general reasoning at all. To verify the hypothesis itself *à posteriori*, that is, to examine whether the facts of any actual case are in accordance with it, is no part of the business of science at all, but of the *application* of science.

In the definition which we have attempted to frame of the science of Political Economy, we have characterized it as essentially an *abstract* science, and its method as the method *à priori*. Such is undoubtedly its character as it has been understood and taught by all its most distinguished teachers. It reasons, and, as we contend, must necessarily reason, from assumptions, not from facts. It is built upon hypotheses, strictly analogous to those which, under the name of definitions, are the foundation of the other abstract sciences. Geometry presupposes an arbitrary definition of a line, "that which has length but not breadth." Just in the same manner does Political Economy presuppose an arbitrary definition of man, as a being who invariably does that by which he may obtain the greatest amount of necessaries, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained in the existing state of knowledge. It is true that this definition of man is not formally prefixed to any work on Political Economy, as the definition of a line is prefixed to Euclid's Elements; and in proportion as by being so prefixed it would be less in danger of being forgotten, we may see ground for regret that this is not done. It is proper that what is assumed in every particular case, should once for all be brought before the mind in its full extent, by being somewhere formally stated as a general maxim. Now, no one who is conversant with systematic treatises on Political Economy will question, that whenever a political economist has shown that, by acting in a particular manner, a labourer may obviously obtain higher wages, a capitalist larger profits, or a landlord higher rent, he concludes, as a matter of course, that they will certainly act in that manner. Political Economy, therefore, reasons from *assumed* premises – from premises which might be totally without foundation in fact, and which are not pretended to be universally in accordance with it. The conclusions of Political Economy, consequently, like those of geometry, are only true, as the common phrase is, *in the abstract*; that is, they are only true under certain suppositions, in which none but general causes – causes common to the *whole class* of cases under consideration – are taken into the account.

This ought not to be denied by the political economist. If he deny it, then, and then only, he places himself in the wrong. The *à priori* method which is laid to his charge, as if his employment of it proved his whole science to be worthless, is, as we shall presently show, the only method by which truth can possibly be attained in any department of the social science. All that is requisite is, that he be on his guard not to ascribe to conclusions which are grounded upon an hypothesis a different kind of certainty from that which really belongs to them. They would be true without qualification, only in a case which is purely imaginary. In proportion as the actual facts recede from the hypothesis, he must allow a corresponding deviation from the strict letter of his conclusion; otherwise it will be true only of things such as he has arbitrarily supposed, not of such things as really exist. That which is true in the abstract, is always true in the concrete with proper *allowances*. When a certain cause really exists, and if left to itself would infallibly produce a certain effect, that same effect, *modified* by all the other concurrent causes, will correctly correspond to the result really produced.

The conclusions of geometry are not strictly true of such lines, angles, and figures, as human hands can construct. But no one, therefore, contends that the conclusions of geometry are of no utility, or that it would be better to shut up Euclid's Elements, and content ourselves with "practice" and "experience."

No mathematician ever thought that his definition of a line corresponded to an actual line. As little did any political economist ever imagine that real men had no object of desire but wealth, or none which would not give way to the slightest motive of a pecuniary kind. But they were justified in assuming this, for the purposes of their argument: because they had to do only with those parts of human conduct which have pecuniary advantage for their direct and principal object; and because, as no two individual cases are exactly alike, no *general* maxim could ever be laid down unless *some* of the circumstances of the particular case were left out of consideration.

But we go farther than to affirm that the method *à priori* is a legitimate mode of philosophical investigation in the moral sciences; we contend that it is the only mode. We affirm that the method *à posteriori*, or that of specific experience, is altogether inefficacious in those sciences, as a means of arriving at any considerable body of valuable truth; though it admits of being usefully applied in aid of the method *à priori*, and even forms an indispensable supplement to it.

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. In chemistry and natural philosophy, we can not only observe what happens under all the combinations of circumstances which nature brings together, but we may also try an indefinite number of new combinations. This we can seldom do in ethical, and scarcely ever in political science. We cannot try forms of government and systems of national policy on a diminutive scale in our laboratories, shaping our experiments as we think they may most conduce to the advancement of knowledge. We therefore study nature under circumstances of great disadvantage in these sciences; being confined to the limited number of experiments which take place (if we may so speak) of their own accord, without any preparation or management of ours; in circumstances, moreover, of great complexity, and never perfectly known to us; and with the far greater part of the processes concealed from our observation.

The consequence of this unavoidable defect in the materials of the induction is, that we can rarely obtain what Bacon has quaintly, but not unaptly, termed an *experimentum crucis*.

In any science which admits of an unlimited range of arbitrary experiments, an *experimentum crucis* may always be obtained. Being able to vary all the circumstances, we can always take effectual means of ascertaining which of them are, and which are not, material. Call the effect B, and let the question be whether the cause A in any way contributes to it. We try an experiment in which all the surrounding circumstances are altered, except A alone: if the effect B is nevertheless produced, A is the cause of it. Or, instead of leaving A, and changing the other circumstances, we leave all the other circumstances and change A: if the effect B in that case does *not* take place, then again A is a necessary condition of its existence. Either of these experiments, if accurately performed, is an *experimentum crucis*; it converts the presumption we had before of the existence of a connection between A and B into proof, by negativing every other hypothesis which would account for the appearances.

But this can seldom be done in the moral sciences, owing to the immense multitude of the influencing circumstances, and our very scanty means of varying the experiment. Even in operating upon an individual mind, which is the case affording greatest room for experimenting, we cannot often obtain a *crucial* experiment. The effect, for example, of a particular circumstance in education, upon the formation of character, may be tried in a variety of cases, but we can hardly ever be certain that any two of those cases differ in all their circumstances except the solitary one of which we wish to estimate the influence. In how much greater a degree must this difficulty exist in the affairs of states, where even the *number* of recorded experiments is so scanty in comparison with the variety and multitude of the circumstances concerned in each. How, for example, can we obtain a crucial experiment on the effect of a restrictive commercial policy upon national wealth? We must find two nations alike in every other respect, or at least possessed, in a degree exactly equal, of everything which conduces to national opulence, and adopting exactly the same policy in all their other affairs, but differing in this only, that one of them adopts a system of commercial restrictions, and the other adopts free trade. This would be a decisive experiment, similar to those which we can almost always obtain in experimental physics. Doubtless this would be the most conclusive evidence of all if we could get it. But let any one consider how infinitely numerous and various are the circumstances which either directly or indirectly do or may influence the amount of the national wealth, and then ask himself what are the probabilities that in the longest revolution of ages two nations will be found, which agree, and can be shown to agree, in all those circumstances except one?

Since, therefore, it is vain to hope that truth can be arrived at, either in Political Economy or in any other department of the social science, while we look at the facts in the concrete, clothed in all the complexity with which nature has surrounded them, and endeavour to elicit a general law by a process of induction from a comparison of details; there remains no other method than the *à priori* one, or that of "abstract speculation."

Although sufficiently ample grounds are not afforded in the field of politics, for a satisfactory induction by a comparison of the effects, the causes may, in all cases, be made the subject of specific experiment. These causes are, laws of human nature, and external circumstances capable of exciting the human will to action. The desires of man, and the nature of the conduct to which they prompt him, are within the reach of our observation. We can also observe what are the objects which excite those desires. The materials of this knowledge every one can principally collect within himself; with reasonable consideration of the differences, of which experience discloses to him the existence, between himself and other people. Knowing therefore accurately the properties of the substances concerned, we may reason with as much certainty as in the most demonstrative parts of physics from any assumed set of circumstances. This will be mere trifling if the assumed circumstances bear no sort of resemblance to any real ones; but if the assumption is correct as far as it goes, and differs from the truth no otherwise than as a part differs from the whole, then the conclusions which are correctly deduced from the assumption constitute *abstract* truth; and when completed by adding or subtracting the effect of the non-calculated circumstances, they are true in the concrete, and may be applied to practice.

Of this character is the science of Political Economy in the writings of its best teachers. To render it perfect as an abstract science, the combinations of circumstances which it assumes, in order to trace their effects, should embody all the circumstances that are common to all cases whatever, and likewise all the circumstances that are common to any important class of cases. The conclusions correctly deduced from these assumptions, would be as true in the abstract as those of mathematics; and would be as near an approximation as abstract truth can ever be, to truth in the concrete.

When the principles of Political Economy are to be applied to a particular case, then it is necessary to take into account all the individual circumstances of that case; not only examining to which of the sets of circumstances contemplated by the abstract science the circumstances of the case in question correspond, but likewise what other circumstances may exist in that case, which not being common to it with any large and strongly-marked class of cases, have not fallen under the cognizance of the science. These circumstances have been called *disturbing causes.* And here only it is that an element of uncertainty enters into the process – an uncertainty inherent in the nature of these complex phenomena, and arising from the impossibility of being quite sure that all the circumstances of the particular case are known to us sufficiently in detail, and that our attention is not unduly diverted from any of them.

This constitutes the only uncertainty of Political Economy; and not of it alone, but of the moral sciences in general. When the disturbing causes are known, the allowance necessary to be made for them detracts in no way from scientific precision, nor constitutes any deviation from the *à priori* method. The disturbing causes are not handed over to be dealt with by mere conjecture. Like *friction* in mechanics, to which they have been often compared, they may at first have been considered merely as a non-assignable deduction to be made by guess from the result given by the general principles of science; but in time many of them are brought within the pale of the abstract science itself, and their effect is found to admit of as accurate an estimation as those more striking effects which they modify. The disturbing causes have their laws, as the causes which are thereby disturbed have theirs; and from the laws of the disturbing causes, the nature and amount of the disturbance may be predicted *à priori*, like the operation of the more general laws which they are said to modify or disturb, but with which they might more properly be said to be concurrent. The effect of the special causes is then to be added to, or subtracted from, the effect of the general ones.

These disturbing causes are sometimes circumstances which operate upon human conduct through the same principle of human nature with which Political Economy is conversant, namely, the desire of wealth, but which are not general enough to be taken into account in the abstract science. Of disturbances of this description every political economist can produce many examples. In other instances the disturbing cause is some other law of human nature. In the latter case it never can fall within the province of Political Economy; it belongs to some other science; and here the mere political economist, he who has studied no science but Political Economy, if he attempt to apply his science to practice, will fail.[1]

As for the other kind of disturbing causes, namely those which operate through the same law of human nature out of which the general principles of the science arise, these might always be brought within the pale of the abstract science if it were worth while; and when we make the necessary allowances for them in practice, if we are doing anything but guess, we are following out the method of the abstract science into minuter details; inserting among its hypotheses a fresh and still more complex combination of circumstances, and so adding *pro hâc vice* a supplementary chapter or appendix, or at least a supplementary theorem, to the abstract science.

Having now shown that the method *à priori* in Political Economy, and in all the other branches of moral science, is the only certain or scientific mode of investigation, and that the *à posteriori* method, or that of specific experience, as a means of arriving at truth, is inapplicable to these subjects, we shall be able to show that the latter method is notwithstanding of great value in the moral sciences; namely, not as a means of discovering truth, but of verifying it, and reducing to the lowest point that uncertainty before alluded to as arising from the complexity of every particular case, and from the difficulty (not to say impossibility) of our being assured *à priori* that we have taken into account all the material circumstances.

If we could be quite certain that we knew all the facts of the particular case, we could derive little additional advantage from specific experience. The causes being given, we may know what will be their effect, without an actual trial of every possible combination; since the causes are human feelings, and outward circumstances fitted to excite them: and, as these for the most part are, or at least might be, familiar to us, we can more surely judge of their combined effect from that familiarity, than from any evidence which can be elicited from the complicated and entangled circumstances of an actual experiment. If the knowledge what are the particular causes operating in any given instance were revealed to us by infallible authority, then, if our abstract science were perfect, we should become prophets. But the causes are not so revealed: they are to be collected by observation; and observation in circumstances of complexity is apt to be imperfect. Some of the causes may lie beyond observation; many are apt to escape it, unless we are on the look-out for them; and it is only the habit of long and accurate observation which can give us so correct a preconception what causes we are likely to find, as shall induce us to look for them in the right quarter. But such is the nature of human understanding, that the very fact of attending with intensity to one part of a thing, has a tendency to withdraw the attention from the other parts. We are consequently in great danger of adverting to a portion only of the causes which are actually at work. And if we are in this predicament, the more accurate our deductions and the more certain our conclusions in the abstract (that is, making abstraction of all circumstances except those which form part of the hypothesis), the less we are likely to suspect that we are in error: for no one could have looked closely into the sources of fallacious thinking without being deeply conscious that the coherence, and neat concatenation of our philosophical systems, is more apt than we are commonly aware to pass with us as evidence of their truth.

We cannot, therefore, too carefully endeavour to verify our theory, by comparing, in the particular cases to which we have access, the results which it would have led us to predict, with most trustworthy accounts we can obtain of those which have been actually realized. The discrepancy between our anticipations and the actual fact is often the only circumstance which would have drawn our attention to some important disturbing cause which we had overlooked. Nay, it often discloses to us errors in thought, still more serious than the omission of what can with any propriety be termed a disturbing cause. It often reveals to us that the basis itself of our whole argument is insufficient; that the data, from which we had reasoned, comprise only a part, and not always the most important part, of the circumstances by which the result is really determined. Such oversights are committed by very good reasoners, and even by a still rarer class, that of good observers. It is a kind of error to which those are peculiarly liable whose views are the largest and most philosophical; for exactly in that ratio are their minds more accustomed to dwell upon those laws, qualities, and tendencies, which are common to large classes of cases, and which belong to all place and all time; while it often happens that circumstances almost peculiar to the particular case or era have a far greater share in governing that one case.

Although, therefore, a philosopher be convinced that no general truths can be attained in the affairs of nations by the *à posteriori* road, it does not the less behove him, according to the measure of his opportunities, to shift and scrutinize the details of every specific experiment. Without this, he may be an excellent professor of abstract science; for a person may be of great use who points out correctly what effects will follow from certain combinations of possible circumstances, in whatever tract of the extensive region of hypothetical cases those combinations may be found. He stands in the same relation to the legislator, as the mere geographer to the practical navigator; telling him the latitude and longitude of all sorts of places, but not how to find whereabouts he himself is sailing. If, however, he does no more than this, he must rest contented to take no share in practical politics; to have no opinion, or to hold it with extreme modesty, on the applications which should be made of his doctrines to existing circumstances.

No one who attempts to lay down propositions for the guidance of mankind, however perfect his scientific acquirements, can dispense with a practical knowledge of the actual modes in which the affairs of the world are carried on, and an extensive personal experience of the actual ideas, feelings, and intellectual and moral tendencies of his own country and of his own age. The true practical statesman is he who combines this experience with a profound knowledge of abstract political philosophy. Either acquirement, without the other, leaves him lame and impotent if he is sensible of the deficiency; renders him obstinate and presumptuous if, as is more probable, he is entirely unconscious of it.

Such, then, are the respective offices and uses of the *à priori* and the *à posteriori* methods – the method of abstract science, and that of specific experiment – as well in Political Economy, as in all the other branches of social philosophy. Truth compels us to express our conviction that whether among those who have written on these subjects, or among those for whose use they wrote, few can be pointed out who have allowed to each of these methods its just value, and systematically kept each to its proper objects and functions. One of the peculiarities of modern times, the separation of theory from practice – of the studies of the closet from the outward business of the world – has given a wrong bias to the ideas and feelings both of the student and of the man of business. Each undervalues that part of the materials of thought with which he is not familiar. The one despises all comprehensive views, the other neglects details. The one draws his notion of the universe from the few objects with which his course of life has happened to render him familiar; the other having got demonstration on his side, and forgetting that it is only a demonstration *nisi* – a proof at all times liable to be set aside by the addition of a single new fact to the hypothesis – denies, instead of examining and sifting, the allegations which are opposed to him. For this he has considerable excuse in the worthlessness of the testimony on which the facts brought forward to invalidate the conclusions of theory usually rest. In these complex matters, men see with their preconceived opinions, not with their eyes: an interested or a passionate man's statistics are of little worth; and a year seldom passes without examples of the astounding falsehoods which large bodies of respectable men will back each other in publishing to the world as facts within their personal knowledge. It is not because a thing is *asserted* to be true, but because in its nature it *may* be true, that a sincere and patient inquirer will feel himself called upon to investigate it. He will use the assertions of opponents not as evidence, but indications leading to evidence; suggestions of the most proper course for his own inquiries.

But while the philosopher and the practical man bandy half-truths with one another, we may seek far without finding one who, placed on a higher eminence of thought, comprehends as a whole what they see only in separate parts; who can make the anticipations of the philosopher guide the observation of the practical man, and the specific experience of the practical man warn the philosopher where something is to be added to his theory.

The most memorable example in modern times of a man who united the spirit of philosophy with the pursuits of active life, and kept wholly clear from the partialities and prejudices both of the student and of the practical statesman, was Turgot; the wonder not only of his age, but of history, for his astonishing combination of the most opposite, and, judging from common experience, almost incompatible excellences.

Though it is impossible to furnish any test by which a speculative thinker, either in Political Economy or in any other branch of social philosophy, may know that he is competent to judge of the application of his principles to the existing condition of his own or any other country, indications may be suggested by the absence of which he may well and surely know that he is not competent. His knowledge must at least enable him to explain and account for what *is*, or he is an insufficient judge of what ought to be. If a political economist, for instance, finds himself puzzled by any recent or present commercial phenomena; if there is any mystery to him in the late or present state of the productive industry of the country, which his knowledge of principle does not enable him to unriddle; he may be sure that something is wanting to render his system of opinions a safe guide in existing circumstances. Either some of the facts which influence the situation of the country and the course of events are not known to him; or, knowing them, he knows not what ought to be their effects. In the latter case his system is imperfect even as an abstract system; it does not enable him to trace correctly all the consequences even of assumed premises. Though he succeed in throwing doubts upon the reality of some of the phenomena which he is required to explain, his task is not yet completed; even then he is called upon to show how the belief, which he deems unfounded, arose; and what is the real nature of the appearances which gave a colour of probability to allegations which examination proves to be untrue.

When the speculative politician has gone through this labour – has gone through it conscientiously, not with the desire of finding his system complete, but of making it so – he may deem himself qualified to apply his principles to the guidance of practice: but he must still continue to exercise the same discipline upon every new combination of facts as it arises; he must make a large allowance for the disturbing influence of unforeseen causes, and must carefully watch the result of every experiment, in order that any residuum of facts which his principles did not lead him to expect, and do not enable him to explain, may become the subject of a fresh analysis, and furnish the occasion for a consequent enlargement or correction of his general views.

The method of the practical philosopher consists, therefore, of two processes; the one analytical, the other synthetical. He must *analyze* the existing state of society into its elements, not dropping and losing any of them by the way. After referring to the experience of individual man to learn the *law* of each of these elements, that is, to learn what are its natural effects, and how much of the effect follows from so much of the cause when not counteracted by any other cause, there remains an operation of *synthesis*; to put all these effects together, and, from what they are separately, to collect what would be the effect of all the causes acting at once. If these various operations could be correctly performed, the result would be prophecy; but as they can be performed only with a certain approximation of correctness, mankind can never predict with absolute certainty, but only with a less or greater degree of probability; according as they are better or worse apprised what the causes are, – have learnt with more or less accuracy from experience the law to which each of those causes, when acting separately, conforms, – and have summed up the aggregate effect more or less carefully.

With all the precautions which have been indicated there will still be some danger of falling into partial views; but we shall at least have taken the best securities against it. All that we can do more, is to endeavour to be impartial critics of our own theories, and to free ourselves, as far as we are able, from that reluctance from which few inquirers are altogether exempt, to admit the reality or relevancy of any facts which they have not previously either taken into, or left a place open for in, their systems.

If indeed every phenomenon was generally the effect of no more than one cause, a knowledge of the law of that cause would, unless there was a logical error in our reasoning, enable us confidently to predict all the circumstances of the phenomenon. We might then, if we had carefully examined our premises and our reasoning, and found no flaw, venture to disbelieve the testimony which might be brought to show that matters had turned out differently from what we should have predicted. If the causes of erroneous conclusions were always patent on the face of the reasonings which lead to them, the human understanding would be a far more trustworthy instrument than it is. But the narrowest examination of the process itself will help us little towards discovering that we have omitted part of the premises which we ought to have taken into our reasoning. Effects are commonly determined by a *concurrence* of causes. If we have overlooked any one cause, we may reason justly from all the others, and only be the further wrong. Our premises will be true, and our reasoning correct, and yet the result of no value in the particular case. There is, therefore, almost always room for a modest doubt as to our practical conclusions. Against false premises and unsound reasoning, a good mental discipline may effectually secure us; but against the danger of *overlooking* something, neither strength of understanding nor intellectual cultivation can be more than a very imperfect protection. A person may be warranted in feeling confident, that whatever he has carefully contemplated with his mind's eye he has seen correctly; but no one can be sure that there is not something in existence which he has not seen at all. He can do no more than satisfy himself that he has seen all that is visible to any other persons who have concerned themselves with the subject. For this purpose he must endeavour to place himself at their point of view, and strive earnestly to see the object as they see it; nor give up the attempt until he has either added the appearance which is floating before them to his own stock of realities, or made out clearly that it is an optical deception.

The principles which we have now stated are by no means alien to common apprehension: they are not absolutely hidden, perhaps, from any one, but are commonly seen through a mist. We might have presented the latter part of them in a phraseology in which they would have seemed the most familiar of truisms: we might have cautioned inquirers against too extensive *generalization*, and reminded them that there are *exceptions* to all rules. Such is the current language of those who distrust comprehensive thinking, without having any clear notion why or where it ought to be distrusted. We have avoided the use of these expressions purposely, because we deem them superficial and inaccurate. The error, when there is error, does *not* arise from generalizing too extensively; that is, from including too wide a range of particular cases in a single proposition. Doubtless, a man often asserts of an entire class what is only true of a part of it; but his error generally consists not in making too wide an assertion, but in making the wrong *kind* of assertion: he predicated an actual result, when he should only have predicated a *tendency* to the result – a power acting with certain intensity in that direction. With regard to *exceptions*; in any tolerably advanced science there is properly no such thing as an exception. What is thought to be an exception to a principle is always some other and distinct principle cutting into the former: some other force which impinges against the first force, and deflects it from its direction. There are not a *law* and an *exception* to that law – the law acting in ninety-nine cases, and the exception in one. There are two laws, each possibly acting in the whole hundred cases, and bringing about a common effect by their conjunct operation. If the force which, being the less conspicuous of the two, is called the disturbing force, prevails sufficiently over the other force in some one case, to constitute that case what is commonly called an exception, the same disturbing force probably acts as a modifying cause in many other cases which no one will call exceptions.

Thus if it were stated to be a law of nature, that all heavy bodies fall to the ground, it would probably be said that the resistance of the atmosphere, which prevents a balloon from falling, constitutes the balloon an exception to that pretended law of nature. But the real law is, that all I heavy bodies *tend* to fall; and to this there is no exception, not even the sun and moon; for even they, as every astronomer knows, tend towards the earth, with force exactly equal to that with which the earth tends towards them. The resistance of the atmosphere might, in the particular case of the balloon, from a misapprehension of what the law of gravitation is, be said to *prevail* over the law; but its disturbing effect is quite as real in every other case, since though it does not prevent, it retards the fall of all bodies whatever. The rule, and the so-called exception, do not divide the cases between them; each of them is a comprehensive rule extending to all cases. To call one of these concurrent principles an exception to the other, is superficial, and contrary to the correct principles of nomenclature and arrangement. An effect of precisely the same kind, and arising from the same cause, ought not to be placed in two different categories, merely as there does or does not exist another cause preponderating over it.

It is only in art, as distinguished from science, that we can with propriety speak of exceptions. Art, the immediate end of which is practice, has nothing to do with causes, except as the means of bringing about effects. However heterogeneous the causes, it carries the effects of them all into one single reckoning, and according as the sum-total is *plus* or *minus*, according as it falls above or below a certain line, Art says, Do this, or Abstain from doing it. The exception does not run by insensible degrees into the rule, like what are called exceptions in science. In a question of practice it frequently happens that a certain thing is either fit to be done, or fit to be altogether abstained from, there being no medium. If, in the majority of cases, it is fit to be done, that is made the rule. When a case subsequently occurs in which the thing ought not to be done, an entirely new leaf is turned over; the rule is now done with, and dismissed: a new train of ideas is introduced, between which and those involved in the rule is a broad line of demarcation; as broad and *tranchant* as the difference between Ay and No. Very possibly, between the last case which comes within the rule and the first of the exception, there is only the difference of a shade: but that shade probably makes the whole interval between acting in one way and in a totally different one. We may, therefore, in talking of art, unobjectionably speak of the *rule* and the *exception*; meaning by the rule, the cases in which there exists a preponderance, however slight, of inducements for acting in a particular way; and by the exception, the cases in which the preponderance is on the contrary side.

Note

1. One of the strongest reasons for drawing the line of separation clearly and broadly between science and art is the following: That the principle of classification in science most conveniently follows the classification of *causes*, while arts must necessarily be classified according to the classification of the *effects*, the production of which is their appropriate end. Now an effect, whether in physics or morals, commonly depends upon a concurrence of causes, and it frequently happens that several of these causes belong to different sciences. Thus in the construction

of engines upon the principles of the science of *mechanics*, it is necessary to bear in mind the *chemical* properties of the material, such as its liability to oxydize; its electrical and magnetic properties, and so forth. From this it follows that although the necessary foundation of all art is science, that is, the knowledge of the properties or laws of the objects upon which, and with which, the art does its work; it is not equally true that every art corresponds to one particular science. Each art presupposes, not one science, but science in general; or, at least, many distinct sciences.

TWO

Objectivity and Understanding in Economics

Max Weber

Max Weber (1864–1920) was born in Erfurt and taught at the universities of Freiburg, Heidelberg, Vienna, and Munich. He is most often regarded as a sociologist, although he was well educated in economics and took an active role in debates about the methodology of economics. He is perhaps best known for his *The Protestant Ethic and the Spirit of Capitalism*, in which he maintains that Calvinism was instrumental in the early development of capitalism; but he made a great many fundamental contributions to our understanding of societies. His methodological writings have also been extremely influential. Reprinted here are excerpts from "'Objectivity' in Social Science and Social Policy," which is probably the best known of his methodological writings.

All serious reflection about the ultimate elements of meaningful human conduct is oriented primarily in terms of the categories "end" and "means." We desire something concretely either "for its own sake" or as a means of achieving something else which is more highly desired. The question of the appropriateness of the means for achieving a given end is undoubtedly accessible to scientific analysis. Inasmuch as we are able to determine (within the present limits of our knowledge) which means for the achievement of a proposed end are appropriate or inappropriate, we can in this way estimate the chances of attaining a certain end by certain available means. In this way we can indirectly criticize the setting of the end itself as practically meaningful (on the basis of the existing historical situation) or as meaningless with reference to existing conditions. Furthermore, when the possibility of attaining a proposed end appears to exist, we can determine (naturally within the limits of our existing knowledge) the consequences which the

---

application of the means to be used will produce in addition to the eventual attainment of the proposed end, as a result of the interdependence of all events. We can then provide the acting person with the ability to weigh and compare the undesirable as over against the desirable consequences of his action. Thus, we can answer the question: what will the attainment of a desired end "cost" in terms of the predictable loss of other values? Since, in the vast majority of cases, every goal that is striven for does "cost" or can "cost" something in this sense, the weighing of the goal in terms of the incidental consequences of the action which realizes it cannot be omitted from the deliberation of persons who act with a sense of responsibility. One of the most important functions of the *technical criticism* which we have been discussing thus far is to make this sort of analysis possible. To apply the results of this analysis in the making of a decision, however, is not a task which science can undertake; it is rather the task of the acting, willing person: he weighs and chooses from among the values involved according to his own conscience and his personal view of the world. Science can make him realize that all action and naturally, according to the circumstances, inaction imply in their consequences the espousal of certain values – and herewith – what is today so willingly overlooked – the rejection of certain others. The act of choice itself is his own responsibility. . . .

The type of social science in which we are interested is an *empirical science* of concrete *reality (Wirklichkeitswissenschaft)*. Our aim is the understanding of the characteristic uniqueness of the reality in which we move. We wish to understand on the one hand the relationships and the cultural significance of individual events in their contemporary manifestations and on the other the causes of their being historically *so* and not *otherwise*. Now, as soon as we attempt to reflect about the way in which life confronts us in immediate concrete situations, it presents an infinite multiplicity of successively and coexistently emerging and disappearing events, both "within" and "outside" ourselves. The absolute infinitude of this multiplicity is seen to remain undiminished even when our attention is focused on a single "object," for instance, a concrete act of exchange, as soon as we seriously attempt an exhaustive description of *all* the individual components of this "individual phenomenon," to say nothing of explaining it causally. All the analysis of infinite reality which the finite human mind can conduct rests on the tacit assumption that only a finite portion of this reality constitutes the object of scientific investigation, and that only it is "important" in the sense of being "worthy of being known." But what are the criteria by which this segment is selected? It has often been thought that the decisive criterion in the cultural sciences, too, was in the last analysis, the "regular" recurrence of certain causal relationships. The "laws" which we are able to perceive in the infinitely manifold stream of events must – according to this conception – contain the scientifically "essential" aspect of reality. As soon as we have shown some causal relationship to be a "law," i.e., if we have shown it to be universally valid by means of comprehensive historical induction or have made it immediately and tangibly plausible according to our subjective experience, a great number of similar cases order themselves under the formula thus attained. Those elements in each individual event which are left unaccounted for by the selection of their elements subsumable under the "law" are considered as scientifically unintegrated residues which will be taken care of in the further perfection of the system of "laws." Alternatively they will be viewed as "accidental" and therefore scientifically unimportant *because* they do not fit into the structure of the "law"; in other words, they are not typical of the event and hence can only be the objects of "idle curiosity." Accordingly, even among the followers of the Historical School we continually find the attitude which declares that the ideal which all the sciences, including the cultural sciences, serve and towards which they should strive even in the remote future is a system of propositions from which reality can be "deduced." As is well known, a leading natural scientist believed that he could designate the (factually unattainable) ideal goal of such a treatment of cultural reality as a sort of "*astronomical*" knowledge. . . .

We have designated as "cultural sciences" those disciplines which analyze the phenomena of life in terms of their cultural significance. The *significance* of a configuration of cultural phenomena and the basis of this significance cannot however be derived and rendered intelligible by a system of analytical laws (*Gesetzesbegriffen*), however perfect it may be, since the significance of cultural events presupposes a *value-orientation* towards these events. The concept of culture is a *value-concept*. Empirical reality becomes "culture" to us because and insofar as we relate it to value ideas. It includes those segments and only those segments of reality which have become significant to us because of this value-relevance. Only a small portion of existing concrete reality is colored by our value-conditioned interest and it alone is significant to us. It is significant because it reveals relationships which are important to us due to their connection with our values. Only because and to the extent that this is the case is it worthwhile for us to know it in its individual features. We cannot discover, however, what is meaningful to us by means of a "presuppositionless" investigation of empirical data. Rather perception of its meaningfulness to us is the presupposition of its becoming an *object* of investigation. Meaningfulness naturally does not coincide with laws as such, and the more general the law the less the coincidence. For the specific meaning which a phenomenon has for us is naturally *not* to be found in those relationships which it shares with many other phenomena. . . .

What is the consequence of all this?

Naturally, it does not imply that the knowledge of *universal* propositions, the construction of abstract concepts, the knowledge of regularities and the attempt to formulate "*laws*" have no scientific justification in the cultural sciences. Quite the contrary, if the causal knowledge of the historians consists of the imputation of concrete effects to concrete causes, a *valid* imputation of any individual effect without the application of "*nomological*" *knowledge* – i.e., the knowledge of recurrent causal sequences – would in general be impossible. Whether a single individual component of a relationship is, in a concrete case, to be assigned causal responsibility for an effect, the causal explanation of which is at issue, can in doubtful cases be determined only by estimating the effects which we *generally* expect from it and from the other components of the same complex which are relevant to the explanation. In other words, the "*adequate*" effects of the causal elements involved must be considered in arriving at any such conclusion. The extent to which the historian (in the widest sense of the word) can perform this imputation in a reasonably certain manner with his imagination sharpened by personal experience and trained in analytic methods and the extent to which he must have recourse to the aid of special disciplines which make it possible, varies with the individual case. Everywhere, however, and hence also in the sphere of complicated economic processes, the more certain and the more comprehensive our general knowledge the greater is the *certainty* of imputation. This proposition is not in the least affected by the fact that even in the case of all so-called "economic laws" without exception, we are concerned here not with "laws" in the narrower exact natural science sense, but with *adequate* causal relationships expressed in rules and with the application of the category of "objective possibility." The establishment of such regularities is not the *end* but rather the *means* of knowledge. It is entirely a question of expediency, to be settled separately for each individual case, whether a regularly recurrent causal relationship of everyday experience should be formulated into a "law." Laws are important and valuable in the exact natural sciences, in the measure that those sciences are *universally valid.* For the knowledge of historical phenomena in their concreteness, the most general laws, because they are most devoid of content, are also the least valuable. The more comprehensive the validity, – or scope – of a term, the more it leads us away from the richness of reality since, in order to include the common elements of the largest possible number of phenomena, it must necessarily be as abstract as possible and hence *devoid* of content. In the cultural sciences, the knowledge of the universal or general is never valuable in itself.

The conclusion which follows from the above is that an "objective" analysis of cultural events, which proceeds according to the thesis that the ideal of science is the reduction of empirical reality of "laws," is meaningless. It is not meaningless, as is often maintained, because cultural or psychic events for instance are "objectively" less governed by laws. It is meaningless for a number of other reasons. Firstly, because the knowledge of social laws is not knowledge of social reality but is rather one of the various aids used by our minds for attaining this end; secondly, because knowledge of *cultural* events is inconceivable except on a basis of the *significance* which the concrete constellations of reality have for us in certain *individual* concrete situations. In *which* sense and in *which* situations this is the case is not revealed to us by any law; it is decided according to the *value-ideas* in the light of which we view "culture" in each individual case. "Culture" is a finite segment of the meaningless infinity of the world process, a segment on which *human beings* confer meaning and significance. This is true for the human being who views a *particular* culture as a mortal enemy and who seeks to "return to nature." He can attain this point of view only after viewing the culture in which he lives from the standpoint of his values, and finding it "too soft." This is the purely logical-formal fact which is involved when we speak of the logically necessary rootedness of all historical entities (*historische Individuen*) in "evaluative ideas." The transcendental presupposition of every *cultural science* lies not in our finding a certain culture or any "culture" in general to be *valuable* but rather in the fact that we are *cultural beings*, endowed with the capacity and the will to take a deliberate attitude towards the world and to lend it *significance*. Whatever this significance may be, it will lead us to judge certain phenomena of human existence in its light and to respond to them as being (positively or negatively) meaningful. Whatever may be the content of this attitude – these phenomena have cultural significance for us and on this significance alone rests its scientific interest. Thus when we speak here of the conditioning of cultural knowledge through *evaluative* ideas (*Wertideen*) (following the terminology of modern, logic), it is done in the hope that we will not be subject to crude misunderstandings such as the opinion that cultural significance should be attributed only to *valuable* phenomena. Prostitution is a *cultural* phenomenon just as much as religion or money. All three are cultural phenomena *only* because and *only* insofar as their existence and the form which they historically assume touch directly or indirectly on our cultural *interests* and arouse our striving for knowledge concerning problems brought into focus by the evaluating ideas which give *significance* to the fragment of reality analyzed by those concepts.

All knowledge of cultural reality, as may be seen, is always knowledge from *particular points of view.* When we require from the historian and social research worker as an elementary presupposition that they distinguish the important from the trivial and that they should have the necessary "point of view" for this distinction, we mean that they must understand how to relate the events of the real world consciously or unconsciously to universal "cultural values" and to select out those relationships which are significant for us. If the notion that those standpoints can be derived from the "facts themselves" continually recurs, it is due to the naive self-deception of the specialist who is unaware that it is due to the evaluative ideas with which he unconsciously approaches his subject matter, that he has selected from an absolute infinity a tiny portion with the study of which he *concerns* himself. In connection with this selection of individual special "aspects" of the event which always and everywhere occurs, consciously or unconsciously, there also occurs that element of cultural-scientific work which is referred to by the often-heard assertion that the "personal" element of a scientific work is what is really valuable in it, and that personality must be expressed in every work if its existence is to be justified. To be sure, without the investigator's evaluative ideas, there would be no principle of selection of subject-matter and no meaningful knowledge of the concrete reality. Just as without the investigator's conviction regarding the significance of particular cultural facts, every attempt to analyze concrete reality is absolutely meaningless, so the direction of his personal belief, the refraction of values in the prism of his mind, gives direction to his work. And the values to which the scientific genius relates the object of his inquiry may determine, i.e., decide the "conception" of a whole epoch, not only concerning what is regarded as "valuable" but also concerning what is significant or insignificant, "important" or "unimportant" in the phenomena.

Accordingly, cultural science in our sense involves "subjective" presuppositions insofar as it concerns itself only with those components of reality which have some relationship, however indirect, to events to which we attach cultural *significance.* Nonetheless, it is entirely *causal* knowledge exactly in the same sense as the knowledge of significant concrete (*individueller*) natural events which have a qualitative character. Among the many confusions which the overreaching tendency of a formal-juristic outlook has brought about in the cultural sciences, there has recently appeared the attempt to "refute" the "materialistic conception of history" by a series of clever but fallacious arguments which state that since all economic life must take place in legally or conventionally *regulated forms*, all economic "development" must take the form of striving for the creation of new *legal* forms. Hence, it is said to be intelligible only through ethical maxims and is on this account essentially different from every type of "natural" development. Accordingly the knowledge of economic development is said to be "teleological" in character. Without wishing to discuss the meaning of the ambiguous term "development," or the logically no less ambiguous term "teleology" in the social sciences, it should be stated that such knowledge need not be "teleological" in the sense assumed by this point of view. The cultural significance of normatively regulated legal *relations* and even norms themselves can undergo fundamental revolutionary changes even under conditions of the formal identity of the prevailing legal norms. Indeed, if one wishes to lose one's self for a moment in phantasies about the future, one might theoretically imagine, let us say, the "socialization of the means of production" unaccompanied by any conscious "striving" towards this result, and without even the disappearance or addition of a single paragraph of our legal code; the statistical frequency of certain legally regulated relationships might be changed fundamentally, and in many cases, even disappear entirely; a great number of legal norms might become *practically* meaningless and their whole cultural significance changed beyond identification. *De lege ferenda* discussions may be justifiably disregarded by the "materialistic conception of history" since its central proposition is the indeed inevitable change in the *significance* of legal institutions. Those who view the painstaking labor of causally understanding historical reality as of secondary importance can disregard it, but it is impossible to supplant it by any type of "teleology." From our viewpoint, "purpose" is the conception of an *effect* which becomes a *cause* of an action. Since we take into account every cause which produces or can produce a significant effect, we also consider this one. Its specific significance consists only in the fact that we not only *observe* human conduct but can and desire to understand it.

Undoubtedly, all evaluative ideas are "subjective." Between the "historical" interest in a family chronicle and that in the development of the greatest conceivable cultural phenomena which were and are common to a nation or to mankind over long epochs, there exists an infinite gradation of "significance" arranged into an order which differs for each of us. And they are, naturally, historically variable in accordance with the character of the culture and the ideas which rule men's minds. But it obviously does not follow from this that research in the cultural sciences can only have results which are "subjective" in the sense that they are *valid* for one person and not for others. Only the degree to which they interest different persons varies. In other words, the choice of the object of investigation and the extent or depth to which this investigation attempts to penetrate into the infinite causal web, are determined by the evaluative ideas which dominate the investigator and his age. In the *method* of investigation, the guiding "point of view" is of great importance for the *construction* of the conceptual scheme which will be used in the investigation. In the mode of their *use*, however, the investigator is obviously bound by the norms of our thought just as much here as elsewhere. For scientific truth is precisely what is *valid* for all who *seek* truth.

However, there emerges from this the meaninglessness of the idea which prevails occasionally even among historians, namely, that the goal of the cultural sciences, however far it may be from realization, is to construct a closed system of concepts, in which reality is synthesized in some sort of *permanently* and *universally* valid classification and from which it can again be deduced. The stream of immeasurable events flows unendingly towards eternity. The cultural problems which move men form themselves ever anew and in different colors, and the boundaries of that area in the infinite stream of concrete events which acquires meaning and significance for us, i.e., which becomes an "historical individual," are constantly subject to change. The intellectual contexts from which it is viewed and scientifically analyzed shift. The points of departure of the cultural sciences remain changeable throughout the limitless future as long as a Chinese ossification of intellectual life does not render mankind incapable of setting new questions to the eternally inexhaustible flow of life. A systematic science of culture, even only in the sense of a definitive, objectively valid, systematic fixation of the problems which it should treat, would be senseless in itself. Such an attempt could only produce a collection of numerous, specifically particularized, heterogeneous and disparate viewpoints in the light of which reality becomes "culture" through being significant in its unique character.

Having now completed this lengthy discussion, we can finally turn to the question which is *methodologically* relevant in the consideration of the "objectivity" of cultural knowledge. The question: what is the logical function and structure of the *concepts* which our science, like all others, uses? Restated with special reference to the decisive problem, the question is: what is the significance of *theory* and theoretical conceptualization (*theoretische Begriffsbildung*) for our knowledge of cultural reality?

Economics was originally – as we have already seen – a "technique," at least in the central focus of its attention. By this we mean that it viewed reality from an at least ostensibly unambiguous and stable practical evaluative standpoint: namely, the increase of the "wealth" of the population. It was on the other hand, from the very beginning, more than a "technique" since it was integrated into the great scheme of the natural law and rationalistic *Weltanschauung* of the eighteenth century. The nature of that *Weltanschauung* with its optimistic faith in the theoretical and practical rationalizability of reality had an important consequence insofar as it *obstructed* the discovery of the *problematic* character of that standpoint which had been assumed as self-evident. As the rational analysis of society arose in close connection with the modern development of natural science, so it remained related to it in its whole method of approach. In the natural sciences, the practical evaluative attitude toward what was immediately and technically useful was closely associated from the very first with the hope, taken over as a heritage of antiquity and further elaborated, of attaining a purely "objective" (i.e., independent of all individual contingencies) monistic knowledge of the totality of reality in a *conceptual* system of metaphysical *validity* and mathematical *form*. It was thought that this hope could be realized by the method of generalizing abstraction and the formulation of laws based on empirical analysis. The natural sciences which were bound to evaluative standpoints, such as clinical medicine and even more what is conventionally called "technology" became purely practical "arts." The values for which they strove, e.g., the health of the patient, the technical perfection of a concrete productive process, etc., were fixed for the time being for all of them. The methods which they used could only consist in the application of the laws formulated by the theoretical disciplines. Every theoretical advance in the construction of these laws was or could also be an advance for the practical disciplines. With the end given, the progressive reduction of concrete practical questions (e.g., a case of illness, a technical problem, etc.) to special cases of generally valid laws, meant that extension of theoretical knowledge was closely associated and identical with the extension of technical-practical possibilities.

When modern biology subsumed those aspects of reality which interest us *historically*, i.e., in all their concreteness, under a universally valid evolutionary principle, which at least had the appearance – but not the actuality – of embracing everything essential about the subject in the scheme of universally valid laws, this seemed to be the final twilight of all evaluative standpoints in all the sciences. For since the so-called historical event was a segment of the totality of reality, since the principle of causality which was the presupposition of all scientific work, seemed to require the analysis of all events into generally valid "laws," and in view of the overwhelming success of the natural sciences which took this idea seriously, it appeared as if there was in general no conceivable meaning of scientific work other than the discovery of the *laws* of events. Only those aspects of phenomena which were involved in the "laws" could be essential from the scientific point of view, and concrete "individual" events could be considered only as "types," i.e., as representative illustrations of laws. An interest in such events in themselves did not seem to be a "scientific" interest.

It is impossible to trace here the important repercussions of this will-to-believe of naturalistic monism in economics. When socialist criticism and the work of the historians were beginning to transform the original evaluative standpoints, the vigorous development of zoological research on one hand and the influence of Hegelian panlogism on the other prevented economics from attaining a clear and full understanding of the relationship between concept and reality. The result, to the extent that we are interested in it, is that despite the powerful resistance to the infiltration of naturalistic dogma due to German idealism since Fichte and the achievement of the German Historical School in law and economics and partly because of the very work of the Historical School, the naturalistic viewpoint in certain decisive problems has not yet been overcome. Among these problems we find the relationship between "theory" and "history," which is still problematic in our discipline.

The "abstract"-theoretical method even today shows unmediated and ostensibly irreconcilable cleavage from empirical-historical research. The proponents of this method recognize in a thoroughly correct way the methodological impossibility of supplanting the historical knowledge of reality by the formulation of laws or, vice versa, of constructing "laws" in the rigorous sense through the mere juxtaposition of historical observations. Now in order to arrive at these laws – for they are certain that science should be directed towards these as its highest goal – they take it to be a fact that we always have a direct awareness of the structure of human actions in all their reality. Hence – so they think – science can make human behavior directly intelligible with axiomatic evidentness and accordingly reveal its laws. The only exact form of knowledge – the formulation of immediately and intuitively *evident* laws – is however at the same time the only one which offers access to events which have not been directly observed. Hence, at least as regards the fundamental phenomena of economic life, the construction of a system of abstract and therefore purely formal propositions analogous to those of the exact natural sciences, is the only means of analyzing and intellectually mastering the complexity of social life. In spite of the fundamental methodological distinction between historical knowledge and the knowledge of "laws" which the creator of the theory drew as the *first* and *only* one, he now claims empirical *validity*, in the sense of the *deducibility* of reality from "laws," for the propositions of abstract theory. It is true that this is not meant in the sense of empirical validity of the abstract economic laws as such, but in the sense that when equally "exact" theories have been constructed for all the other relevant factors, all these abstract theories together must contain the true reality of the object – i.e., whatever is worthwhile knowing about it. Exact economic theory deals with the operation of *one* psychic motive, the other theories have as their task the formulation of the behavior of all the other motives into similar sorts of propositions enjoying hypothetical validity. Accordingly, the fantastic claim has occasionally been made for economic theories – e.g., the abstract theories of price, interest, rent, etc., – that they can, by ostensibly following the analogy of physical science propositions, be validly applied to the derivation of quantitatively stated conclusions from given real premises, since given the ends, economic behavior with respect to means is unambiguously "determined." This claim fails to observe that in order to be able to reach this result even in the simplest case, the totality of the existing historical reality including every one of its causal relationships must be assumed as "given" and presupposed as known. But if *this* type of knowledge were accessible to the finite mind of man, abstract theory would have no cognitive value whatsoever. The naturalistic prejudice that every concept in the cultural sciences should be similar to those in the exact natural sciences has led in consequence to the misunderstanding of the meaning of this theoretical construction (*theoretische Gedankengebilde*). It has been believed that it is a matter of the psychological isolation of a specific "impulse," the acquisitive impulse, or of the isolated study of a specific maxim of human conduct, the so-called economic principle. Abstract theory purported to be based on psychological *axioms* and as a result historians have called for an *empirical* psychology in order to show the invalidity of those axioms and to derive the course of economic events from psychological principles. We do not wish at this point to enter into a detailed criticism of the belief in the significance of a – still to be created – systematic science of "social psychology" as the future foundation of the cultural sciences, and particularly of social economics, Indeed, the partly brilliant attempts which have been made hitherto to interpret economic phenomena psychologically, show in any case that the procedure does not begin with the analysis of psychological qualities, moving then to the analysis of social institutions, but that, on the contrary, insight into the psychological preconditions and consequences of institutions presupposes a precise knowledge of the latter and the scientific analysis of their structure. In concrete cases, psychological analysis can contribute then an extremely valuable deepening of the knowledge of the historical cultural *conditioning* and cultural *significance* of institutions. The interesting aspect of the psychic attitude of a person in a social situation is specifically particularized in each case, according to the special cultural significance of the situation in question. It is a question of an extremely heterogeneous and highly concrete structure of psychic motives and influences. Social-psychological research involves the study of various very disparate *individual* types of cultural elements with reference to their interpretability by our empathic understanding. Through social-psychological research, with the knowledge of individual institutions as a point of departure, we will learn increasingly how to understand institutions in a psychological way. We will not however deduce the institutions from psychological laws or explain them by elementary psychological phenomena.

Thus, the far-flung polemic, which centered on the question of the psychological justification of abstract theoretical propositions, on the scope of the "acquisitive impulse" and the "economic principle," etc., turns out to have been fruitless.

In the establishment of the propositions of abstract theory, it is only apparently a matter of "deductions" from fundamental psychological motives. Actually, the former are a special case of a kind of concept-construction which is peculiar and to a certain extent, indispensable, to the cultural sciences. It it worthwhile at this point to describe it in further detail since we can thereby approach more closely the fundamental question of the significance of theory in the social sciences. Therewith we leave undiscussed, once and for all, whether *the* particular analytical concepts which we cite or to which we allude as illustrations, correspond to the purposes they are to serve, i.e., whether in fact they are well-adapted. The question as to how far, for example, contemporary "abstract theory" should be further elaborated, is ultimately also a question of the strategy of science, which must, however concern itself with other problems as well. Even the "theory of marginal utility" is subsumable under a "law of marginal utility."

We have in abstract economic theory an illustration of those synthetic constructs which have been designated as "*ideas*" of historical phenomena. It offers us an ideal picture of events on the commodity-market under conditions of a society organized on the principles of an exchange economy, free competition and rigorously rational conduct. This conceptual pattern brings together certain relationships and events of historical life into a complex, which is conceived as an internally consistent system. Substantively, this construct in itself is like a *utopia* which has been arrived at by the analytical accentuation of certain elements of reality. Its relationship to the empirical data consists solely in the fact that where market-conditioned relationships of the type referred to by the abstract construct are discovered or suspected to exist in reality to some extent, we can make the *characteristic* features of this relationship pragmatically *clear* and *understandable* by reference to an *ideal-type.* This procedure can be indispensable for heuristic as well as expository purposes. The ideal typical concept will help to develop our skill in imputation in *research:* it *is* no "hypothesis" but it offers guidance to the construction of hypotheses. It is not a *description* of reality but it aims to give unambiguous means of expression to such a description. It is thus the "idea" of the *historically* given modern society, based on an exchange economy, which is developed for us by quite the same logical principles as are used in constructing the idea of the medieval "city economy" as a "genetic" concept. When we do this, we construct the concept "city economy" not as an average of the economic structures actually existing in all the cities observed but as an *ideal-type.* An ideal type is formed by the one-sided *accentuation* of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent *concrete individual* phenomena, which are arranged according to those one-sidedly emphasized viewpoints into a unified *analytical* construct (*Gedankenbild*). In its conceptual purity, this mental construct (*Gedankenbild*) cannot be found empirically anywhere in reality. It is a *utopia.* Historical research faces the task of determining in each individual case, the extent to which this ideal construct approximates to or diverges from reality, to what extent for example, the economic structure of a certain city is to be classified as a "city-economy." When carefully applied, those concepts are particularly useful in research and exposition. In very much the same way one can work the "idea" of "handicraft" into a Utopia by arranging certain traits, actually found in an unclear, confused state in the industrial enterprises of the most diverse epochs and countries, into a consistent ideal-construct by an accentuation of their essential tendencies. This ideal-type is then related to the idea (*Gedankenausdruck*) which one finds expressed there. One can further delineate a society in which all branches of economic and even intellectual activity are governed by maxims which appear to be applications of the same principle which characterizes the ideal-typical "handicraft" system. Furthermore, one can juxtapose alongside the ideal typical "handicraft" system the antithesis of a correspondingly ideal-typical capitalistic productive system, which has been abstracted out of certain features of modern large scale industry. On the basis of this, one can delineate the Utopia of a "capitalistic" culture, i.e., one in which the governing principle is the investment of private capital. This procedure would accentuate certain individual concretely diverse traits of modern material and intellectual culture in its unique aspects into an ideal construct which from our point of view would be completely self-consistent. This would then be the delineation of an "*idea*" of *capitalistic culture.* We must disregard for the moment whether and how this procedure could be carried out. It is possible, or rather, it must be accepted as certain that numerous, indeed a very great many, Utopias of this sort can be worked out, of which *none* is like another, and *none* of which can be observed in empirical reality as an actually existing economic system, but *each* of which however claims that it is a representation of the "idea" of capitalistic culture. *Each* of these can claim to be a representation of the "idea" of capitalistic culture to the extent that it has really taken certain traits, meaningful in their essential features, from the empirical reality of our culture and brought them together into a unified ideal-construct. For those phenomena which interest us as cultural phenomena are interesting to us with respect to very different kinds of evaluative ideas to which we relate them. Inasmuch as the "points of view" from which they can become significant for us are very diverse, the most varied criteria can be applied to the selection of the traits which are to enter into the construction of an ideal-typical view of a particular culture.

THREE

The Nature and Significance of Economic Science

Lionel Robbins

Lionel Robbins (1898–1984) was educated at University College, London, and at the London School of Economics. He taught at the London School of Economics, New College, Oxford, and at the University of London. The author of many books on economics and on economic policy, Robbins was made a life peer in 1959. His *An Essay on the Nature and Significance of Economic Science*, from which excerpts are reprinted here, was immediately recognized as a classic and has been very influential.

Chapter I: The Subject Matter of Economics

. . .3. But where, then, are we to turn? The position is by no means hopeless. Our critical examination of the "materialist" definition has brought us to a point from which it is possible to proceed forthwith to formulate a definition which shall be immune from all these strictures.

Let us turn back to the simplest case in which we found this definition inappropriate – the case of isolated man dividing his time between the production of real income and the enjoyment of leisure. We have just seen that such a division may legitimately be said to have an economic aspect. Wherein does this aspect consist?

The answer is to be found in the formulation of the exact conditions which make such division necessary. They are four. In the first place, isolated man wants both real income and leisure. Secondly, he has not enough of either fully to satisfy his want of each. Thirdly, he can spend his time in augmenting his real income or he can spend it in taking more leisure. Fourthly, it may be presumed that, save in most exceptional cases, his want for the different constituents of real income and leisure will be different. Therefore he has to

---

Excerpted from *An Essay on the Nature and Significance of Economic Science*, 2d ed., by Lionel Robbins. London: Macmillan, 1935.

choose. He has to economise. The disposition of his time and his resources has a relationship to his system of wants. It has an economic aspect.

This example is typical of the whole field of economic studies. From the point of view of the economist, the conditions of human existence exhibit four fundamental characteristics. The ends are various. The time and the means for achieving these ends are limited and capable of alternative application. At the same time the ends have different importance. Here we are, sentient creatures with bundles of desires and aspirations, with masses of instinctive tendencies all urging us in different ways to action. But the time in which these tendencies can be expressed is limited. The external world does not offer full opportunities for their complete achievement. Life is short. Nature is niggardly. Our fellows have other objectives. Yet we can use our lives for doing different things, our materials and the services of others for achieving different objectives.

Now *by itself* the multiplicity of ends has no necessary interest for the economist. If I want to do two things, and I have ample time and ample means with which to do them, and I do not want the time or the means for anything else, then my conduct assumes none of those forms which are the subject of economic science. Nirvana is not necessarily single bliss. It is merely the complete satisfaction of *all* requirements.

Nor is the mere limitation of means *by itself* Sufficient to give rise to economic phenomena. If means of satisfaction have no alternative use, then they may be scarce, but they cannot be economised The Manna which fell from heaven may have been scarce, but, if it was impossible to exchange it for something else or to postpone its use,[1] it was not the object of any activity with an economic aspect.

Nor again is the alternative applicability of scarce means a complete condition of the existence of the kind of phenomena we are analysing. If the economic subject has two ends and one means of satisfying them, and the two ends are of equal importance, his position will be like the position of the ass in the fable, paralysed halfway between the two equally attractive bundles of hay.[2]

But when time and the means for achieving ends are limited *and* capable of alternative application, *and* the ends are capable of being distinguished in order of importance, then behaviour necessarily assumes the form of choice. Every act which involves time and scarce means for the achievement of one end involves the relinquishment of their use for the achievement of another. It has an economic aspect.[3] If I want bread and sleep, and in the time at my disposal I cannot have all I want of both, then some part of my wants of bread and sleep must go unsatisfied. If, in a limited lifetime, I would wish to be both a philosopher and a mathematician, but my rate of acquisition of knowledge is such that I cannot do both completely, then some part of my wish for philosophical or mathematical competence or both must be relinquished.

Now not all the means for achieving human ends are limited. There are things in the external world which are present in such comparative abundance that the use of particular units for one thing does not involve going without other units for others. The air which we breathe, for instance, is such a "free" commodity. Save in very special circumstances, the fact that we need air imposes no sacrifice of time or resources. The loss of one cubic foot of air implies no sacrifice of alternatives. Units of air have no specific significance for conduct. And it is conceivable that living creatures might exist whose "ends" were so limited that all goods for them were "free" goods, that no goods had specific significance.

But, in general, human activity with its multiplicity of objectives has not this independence of time or specific resources. The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. The services which others put at our disposal are limited. The material means of achieving ends are limited. We have been turned out of Paradise. We have neither eternal life nor unlimited means of gratification. Everywhere we turn, if we choose one thing we must relinquish others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy ends of varying importance is an almost ubiquitous condition of human behaviour.[4]

Here, then, is the unity of subject of Economic Science, the forms assumed by human behaviour in disposing of scarce means. The examples we have discussed already harmonise perfectly with this conception. Both the services of cooks and the services of opera dancers are limited in relation to demand and can be put to alternative uses. The theory of wages in its entirety is covered by our present definition. So, too, is the political economy of war. The waging of war necessarily involves the withdrawal of scarce goods and services from other uses, if it is to be satisfactorily achieved. It has therefore an economic aspect. The economist studies the disposal of scarce means. He is interested in the way different degrees of scarcity of different goods give rise to different ratios of valuation between them, and he is interested in the way in which changes in conditions of scarcity, whether coming from changes in ends or changes in means – from the demand side or the supply side – affect these ratios. Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses.[5] . . .

Chapter IV: The Nature of Economic Generalisations

1. We have now sufficiently discussed the subject-matter of Economics and the fundamental conceptions associated therewith. But we have not yet discussed the nature of the generalisations whereby these conceptions are related. We have not yet discussed *the nature and derivation of economic laws.* This, therefore, is the purpose of the present chapter. When it is completed we shall be in a position to proceed to our second main task – investigation of the limitations and significance of this system of generalisations.

2. It is the object of this essay to arrive at conclusions which are based on the inspection of Economic Science as it actually exists. Its aim is not to discover how Economics should be pursued – that controversy, although we shall have occasion to refer to it *en passant*,[6] may be regarded as settled as between reasonable people – but rather what significance is to be attached to the results which it has already achieved. It will be convenient, therefore, at the outset of our investigations, if, instead of attempting to derive the nature of economic generalisations from the pure categories of our subject-matter,[7] we proceed rather by examining specimens drawn from the existing body of analysis.

The most fundamental propositions of economic analysis are the propositions of the general theory of value. No matter what particular "school" is in question, no matter what arrangement of subject-matter is adopted, the body of propositions explaining the nature and the determination of the relation between given goods of the first order will be found to have a pivotal position in the whole system. It would be premature to say that the theory of this part of the subject is complete. But it is clear that enough has been done to warrant our taking the central propositions as established. We may proceed, therefore, to inquire on what their validity depends.

It should not be necessary to spend much time showing that it cannot rest upon a mere appeal to "History". The frequent concomitance of certain phenomena in time may suggest a problem to be solved. It cannot by itself be taken to imply a definite causal relationship. It might be shown that, whenever the conditions postulated in any of the simple corollaries of the theory of value have actually existed, the consequences deduced have actually been observed to follow. Thus, whenever the fixing of prices in relatively free markets has taken place it has been followed either by evasion or by the kind of distributive chaos which we associate with the food queues of the late war or the French or Russian Revolutions.[8] But this would not prove that the phenomena in question were causally connected in any intimate sense. Nor would it afford any safe ground for predictions with regard to their future relationship. In the absence of rational grounds for supposing intimate connection, there would be no sufficient reason for supposing that history "would repeat itself." For if there is one thing which *is* shown by history, not less than by elementary logic, it is that historical induction, unaided by the analytical judgment, is the worst possible basis of prophecy.[9] "History shows", commences the bore at the club, and we resign ourselves to the prediction of the improbable. It is one of the great merits of the modern philosophy of history that it has repudiated all claims of this sort, and indeed makes it the *fundamentum divisionis* between history and natural science that history does not proceed by way of generalising abstraction.[10]

It is equally clear that our belief does not rest upon the results of controlled experiment. It is perfectly true that the particular case just mentioned has on more than one occasion been exemplified by the results of government intervention carried out under conditions which might be held to bear some resemblance to the conditions of controlled experiment. But it would be very superficial to suppose that the results of these "experiments" can be held to justify a proposition of such wide applicability, let alone the central propositions of the general theory of value. Certainly it would be a very fragile body of economic generalisations which could be erected on a basis of this sort. Yet, in fact, our belief in these propositions is as complete as belief based upon any number of controlled experiments.

But on what, then, does it depend?

It does not require much knowledge of modern economic analysis to realise that the foundation of the theory of value is the assumption that the different things that the individual wants to do have a different importance to him, and can be arranged therefore in a certain order. This notion can be expressed in various ways and with varying degrees of precision, from the simple want systems of Menger and the early Austrians to the more refined scales of relative valuations of Wicksteed and Schönfeld and the indifference systems of Pareto and Messrs. Hicks and Allen. But in the last analysis it reduces to this, that we can judge whether different possible experiences are of equivalent or greater or less importance to us. From this elementary fact of experience we can derive the idea of the substitutability of different goods, of the demand for one good in terms of another, of an equilibrium distribution of goods between different uses, of equilibrium of exchange and of the formation of prices. As we pass from the description of the behaviour of the single individual to the discussion of markets we naturally make other subsidiary assumptions – there are two individuals or many, the supply is in the hands of a monopoly or of a multiplicity of sellers, the individuals in one part of the market know or do not know what is going on in other parts of the market, the legal framework of the market prohibits this or that mode of acquisition of exchange, and so on. We assume, too, a given initial distribution of property.[11] But always the main underlying assumption is the assumption of the schemes of valuation of the different economic subjects. But this, we have seen already, is really an assumption of one of the conditions which must be present if there is to be economic activity at all. It is an essential constituent of our conception of conduct with an economic aspect.

The propositions so far mentioned all relate to the theory of the valuation of given goods. In the elementary theory of value and exchange no inquiry is made into the conditions of continuous production. If we assume that production takes place, a new set of problems arises, necessitating new principles of explanation. We are confronted, *e.g.*, with the problem of explaining the relation between the value of the products and the value of the factors which produced them – the so-called problem of imputation. What is the sanction here for the solutions which have been put forward?

As is well known, the main principle of explanation, supplementary to the principles of subjective valuation assumed in the narrower theory of value and exchange, is the principle sometimes described as the Law of Diminishing Returns. Now the Law of Diminishing Returns is simply one way of putting the obvious fact that different factors of production are imperfect substitutes for one another. If you increase the amount of labour without increasing the amount of land the product will increase, but it will not increase proportionately. To secure a doubling of the product, if you do not double both land and labour, you have to more than double either one of the factors. This is obvious. If it were not so, then all the corn in the world could be produced from one acre of land. It follows, too, from considerations more intimately connected with our fundamental conceptions. A class of scarce factors is to be defined as consisting of those factors which are perfect substitutes. That is to say, differences in factors is to be defined essentially as imperfect substitutability. The Law of Diminishing Returns, therefore, follows from the assumption that there is more than one class of scarce factors of production.[12] The supplementary principle that, within limits, returns may increase, follows equally directly from the assumption that factors are relatively indivisible. On the basis of these principles and with the aid of subsidiary assumptions of the kind already mentioned (the nature of markets and the legal framework of production, etc.), it is possible to build up a theory of equilibrium of production.[13]

Let us turn to more dynamic considerations. The theory of profits, to use the word in the rather restricted sense in which it has come to be used in recent theory, is essentially an analysis of the effects of uncertainty with regard to the future availability of scarce goods and scarce factors. We live in a world in which, not only are the things that we want scarce, but their exact occurrence is a matter of doubt and conjecture. In planning for the future we have to choose, not between certainties, but rather between a range of estimated probabilities. It is clear that the nature of this range itself may vary, and accordingly there must arise not only relative valuation of the different kinds of uncertainties between themselves, but also of different ranges of uncertainty similarly compared. From such concepts may be deduced many of the most complicated propositions of the theory of economic dynamics.[14]

And so we could go on. We could show how the use of money can be deduced from the existence of indirect exchange and how the demand for money can be deduced from the existence of the same uncertainties that we have just examined.[15] We could examine the propositions of the theory of capital and interest, and reduce them to elementary concepts of the type we have been here discussing. But it is unnecessary to prolong the discussion further. The examples we have already examined should be sufficient to establish the solution for which we are seeking. The propositions of economic theory, like all scientific theory, are obviously deductions from a series of postulates. And the chief of these postulates are all assumptions involving in some way simple and indisputable facts of experience relating to the way in which the scarcity of goods which is the subject-matter of our science actually shows itself in the world of reality. The main postulate of the theory of value is the fact that individuals can arrange their preferences in an order, and in fact do so. The main postulate of the theory of production is the fact that there are more than one factor of production. The main postulate of the theory of dynamics is the fact that we are not certain regarding future scarcities. These are not postulates the existence of whose counterpart in reality admits of extensive dispute once their nature is fully realised. We do not need controlled experiments to establish their validity: they are so much the stuff of our everyday experience that they have only to be stated to be recognised as obvious. Indeed, the danger is that they may be thought to be so obvious that nothing significant can be derived from their further examination. Yet in fact it is on postulates of this sort that the complicated theorems of advanced analysis ultimately depend. And it is from the existence of the conditions they assume that the general applicability of the broader propositions of economic science is derived.

3. Now of course it is true, as we have already seen, that the development of the more complicated applications of these propositions involves the use of a great multitude of subsidiary postulates regarding the condition

of markets, the number of parties to the exchange, the state of the law, the *minimum sensible* of buyers and sellers, and so on and so forth. The truth of the deductions from this structure depends, as always, on their logical consistency. Their applicability to the interpretation of any particular situation depends upon the existence in that situation of the elements postulated. Whether the theory of competition or of monopoly is applicable to a given situation is a matter for inquiry. As in the applications of the broad principles of the natural sciences, so in the application of economic principles we must be careful to enquire concerning the nature of our material. It is not assumed that any of the many possible forms of competitive or monopolistic conditions *must* necessarily always exist. But while it is important to realise how many are the subsidiary assumptions which necessarily arise as our theory becomes more and more complicated, it is equally important to realise how widely applicable are the main assumptions on which it rests. As we have seen, the chief of them are applicable whenever and wherever the conditions which give rise to economic phenomena are present.

Considerations of this sort, it may be urged, should enable us easily to detect the fallacy implicit in a view which has played a great role in continental discussions. It has sometimes been asserted that the generalisations of Economics are essentially "historico-relative" in character, that their validity is limited to certain historical conditions, and that outside these they have no relevance to the analysis of social phenomena. This view is a dangerous misapprehension. It can be given plausibility only by a distortion of the use of words so complete as to be utterly misleading. It is quite true that in order fruitfully to apply the more general propositions of Economics, it is important to supplement them with a series of subsidiary postulates drawn from the examination of what may often be legitimately designated historico-relative material. It is certain that unless this is done bad mistakes are likely to be made. But it is not true that the main assumptions are historico-relative *in the same sense.* It is true that they are based upon experience, that they refer to reality. But it is experience of so wide a degree of generality as to place them in quite a different class from the more properly designated historico-relative assumptions. No one will really question the universal applicability of such assumptions as the existence of scales of relative valuation, or of different factors of production, or of different degrees of uncertainty regarding the future, even though there may be room for dispute as to the best mode of describing their exact logical status. And no one who has really examined the kind of deductions which can be drawn from such assumptions can doubt the utility of starting from this plane. It is only failure to realise this, and a too exclusive preoccupation with the subsidiary assumptions, which can lend any countenance to the view that the laws of Economics are limited to certain conditions of time and space, that they are purely historical in character, and so on. If such views are interpreted to mean merely that we must realise that the applications of general analysis involve a host of subsidiary assumptions of a less general nature, that before we apply our general theory to the interpretation of a particular situation we must be sure of the facts – well and good. Any teacher who has watched good students over-intoxicated with the excitement of pure theory will agree. It may even be conceded that at times there may have been this degree of justification in the criticisms of the classical economists by the better sort of historian. But if, as in the history of the great methodological controversies has notoriously been the case, they are interpreted to mean that the broad conclusions springing from general analysis are as limited as their particular applications – that the generalisations of Political Economy were applicable only to the state of England in the early part of the reign of Queen Victoria, and such-like contentions – then it is clearly utterly misleading. There is perhaps a sense in which it is true to say that *all* scientific knowledge is historico-relative. Perhaps in some other existence it would all be irrelevant. But if this is so, then we need a new term to designate what is usually called historico-relative. So with that body of knowledge which is general economics. If it is historico-relative, then a new term is needed to describe what we know as historico-relative studies.

Stated this way, surely the case for the point of view underlying the so-called "orthodox" conception of the science since the time of Senior and Cairnes is overwhelmingly convincing. It is difficult to see why there should have been such fuss, why anybody should have thought it worth while calling the whole position in question. And, of course, if we examine the actual history of the controversy it becomes abundantly clear that the case for the attack was not primarily scientific and philosophical at all. It may have been the case that from time to time a sensitive historian was outraged by the crudities of some very second-rate economist – more probably by some business man or politician repeating at second-hand what he thought the economists had said. It may have been the case sometimes that a pure logician has been offended by an incautious use of philosophical terms on the part of an economist, anxious to vindicate a body of knowledge which he knows to be true and important. But in the main the attacks have not come from these quarters. Rather they have been *political* in nature. They have come from men with an axe to grind – from men who wished to pursue courses which the acknowledgment of law in the economic sphere would have suggested to be unwise. This was certainly the case with the majority of the leaders of the younger Historical School,[16] who were the spearhead of the attack on international liberalism in the Bismarckian era. It is equally the case to-day with the lesser schools which adopt a similar attitude. The only difference between Institutionalism and *Historismus* is that *Historismus* is much more interesting.

4. If the argument which has been developed above is correct, economic analysis turns out to be as Fetter has emphasised,[17] the elucidation of the implications of the necessity of choice in various assumed circumstances. In pure Mechanics we explore the implication of the existence of certain given properties of bodies. In pure Economics we examine the implication of the existence of scarce means with alternative uses. As we have seen, the assumption of relative valuations is the foundation of all subsequent complications.

It is sometimes thought, even at the present day, that this notion of relative valuation depends upon the validity of particular psychological doctrines. The borderlands of Economics are the happy hunting-ground of minds averse to the effort of exact thought, and, in these ambiguous regions, in recent years, endless time has been devoted to attacks on the alleged psychological assumptions of Economic Science. Psychology, it is said, advances very rapidly. If, therefore, Economics rests upon particular psychological doctrines, there is no task more ready to hand than every five years or so to write sharp polemics showing that, since psychology has changed its fashion, Economics needs "rewriting from the foundations upwards". As might be expected, the opportunity has not been neglected. Professional economists, absorbed in the exciting task of discovering new truth, have usually disdained to reply: and the lay public, ever anxious to escape a necessity of recognising the implications of choice in a world of scarcity, has allowed itself to be bamboozled into believing that matters, which are in fact as little dependent oh the truth of fashionable psychology as the multiplication table, are still open questions on which the enlightened man, who, of course, is nothing if not a psychologist, must be willing to suspend judgment.

Unfortunately, in the past, incautious utterances on the part of economists themselves have sometimes afforded a pretext for these strictures. It is well known that certain of the founders of the modern subjective theory of value did in fact claim the authority of the doctrines of psychological hedonism as sanctions for their propositions. This was not true of the Austrians. From the beginning the Mengerian tables were constructed in terms which begged no psychological questions.[18] Böhm-Bawerk explicitly repudiated any affiliation with psychological hedonism; indeed, he went to infinite pains to avoid this kind of misconception.[19] But the names of Gossen and Jevons and Edgeworth, to say nothing of their English followers, are a sufficient reminder of a line of really competent economists who did make pretensions of this sort. Gossen's *Entwicklung der Gesetze des menschlichen Verkehrs* certainly invokes hedonistic postulates. Jevons in his *Theory of Political Economy* prefaces his theory of utility and exchange with a theory of pleasure and pain. Edgeworth commences his *Mathematical Psychics* with a section which urges the conception of "man as a pleasure machine".[20] Attempts have even been made to exhibit the law of diminishing marginal utility as a special case of the Weber-Fechner Law.[21]

But it is fundamentally important to distinguish between the actual practice of economists, and the logic which it implies, and their occasional *ex post facto* apologia. It is just this distinction which the critics of Economic Science fail to make. They inspect with supererogatory zeal the external façade, but they shrink from the intellectual labour of examining the inner structure. Nor do they trouble to acquaint themselves with the more recent formulations of the theory they are attacking. No doubt this has strategic advantages, for, in polemics of this kind, honest misconception is an excellent spur to effective rhetoric; and no one who was acquainted with recent value theory could honestly continue to argue that it has any essential connection with psychological hedonism, or for that matter with any other brand of *Fach-Psychologie*. If the psychological critics of Economics had troubled to do these things they would speedily have perceived that the hedonistic trimmings of the works of Jevons and his followers were incidental to the main structure of a theory which – as the parallel development in Vienna showed – is capable of being set out and defended in absolutely non-hedonistic terms. As we have seen already, all that is assumed in the idea of the scales of valuation is that different goods have different uses and that these different uses have different significances for action, such that in a given situation one use will be preferred before another and one good before another. Why the human animal attaches particular values in this sense to particular things, is a question which we do not discuss. That is quite properly a question for psychologists or perhaps even physiologists. All that we need to assume as economists is the obvious fact that different possibilities offer different incentives, and that these incentives can be arranged in order of their intensity. The various theorems which may be derived from this fundamental conception are unquestionably capable of explaining a manifold of social activity incapable of explanation by any other technique. But they do this, not by assuming some particular psychology, but by regarding the things which psychology studies as the data of their own deductions. Here, as so often, the founders of Economic Science constructed something more universal in its application than anything that they themselves claimed.

But now the question arises how far even this procedure is legitimate. It should be clear from all that has been said already that although it is not true that the propositions of analytical economics rest upon any particular psychology, yet they do most unquestionably involve elements which are of a psychological – or perhaps better said a psychical – nature. This, indeed, is explicitly recognised in the name by which they are sometimes known – the subjective or psychological theory of value; and, as we have seen, it is clear that the foundation of this theory is a psychical fact, the valuations of the individual. In recent years, however, partly as a result of the influence of Behaviourism, partly as a result of a desire to secure the maximum possible austerity in analytical exposition, there have arisen voices urging that this framework of subjectivity should be discarded. Scientific method, it is urged, demands that we should leave out of account anything which is incapable of direct observation. We may take account of demand as it shows itself in observable behaviour in the market. But beyond this we may not go. Valuation is a subjective process. We cannot *observe* valuation. It is therefore out of place in a scientific explanation. Our theoretical constructions must assume observable data. Such, for instance, is the atttitude of Professor Cassel,[22] and there are passages in the later work of Pareto[23] which permit of a similar interpretation. It is an attitude which is very frequent among those economists who have come under the influence of Behaviourist psychology or who are terrified of attack from exponents of this queer cult.

At first sight this seems very plausible. The argument that we should do nothing that is not done in the physical sciences is very seductive. But it is doubtful whether it is really justified. After all, our business is to explain certain aspects of conduct. And it is very questionable whether this can be done in terms which involve no psychical element. It is quite certain that whether it be pleasing or no to the desire for the maximum austerity, we do in fact *understand* terms such as choice, indifference, preference, and the like in terms of inner experience. The idea of an end, which is fundamental to our conception of the economic, is not possible to define in terms of external behaviour only. If we are to explain the relationships which arise from the existence of a scarcity of means in relation to a multiplicity of ends, surely at least one-half of the equation, as it were, must be psychical in character.

Such considerations would be decisive so long as it were taken for granted that the definition of the subject-matter of Economics suggested in this essay was correct. But it might be urged that they were simply an argument for rejecting that definition and substituting one relating only to "objective", observable matters, market prices, ratios of exchange, and so on. This is clearly what is implied by Professor Cassel's procedure – the celebrated *Ausschaltung der Wertlehre.*

But even if we restrict the object of Economics to the explanation of such observable things as prices, we shall find that in fact it is impossible to explain them unless we invoke elements of a subjective or psychological nature. It is surely clear, as soon as it is stated specifically, that the most elementary process of price determination must depend *inter alia* upon what people think is going to happen to prices in the future. The demand functions which Professor Cassel thinks enable us to dispense with any subjective elements, must be conceived not merely as relating to prices which prevail now, or which might prevail, on present markets, but also as relating to a whole series of prices which people expect to prevail in the future. It is obvious that what people expect to happen in the future is not susceptible of observation by purely behaviourist methods. Yet, as Professor Knight and others have shown, it is absolutely essential to take such anticipations into account if we are to understand at all the mechanics of economic change. It is essential for a thorough explanation of competitive prices. It is indispensable for the most superficial explanation of monopolistic prices. It is quite easy to exhibit such anticipations as part of a general system of scales of preference.[24] But if we suppose that such a system takes account of observable data only we deceive ourselves. How can we observe what a man thinks is going to happen?

It follows, then, that if we are to do our job as economists, if we are to provide a sufficient explanation of matters which every definition of our subject-matter necessarily covers, we must include psychological elements. They cannot be left out if our explanation is to be adequate. It seems, indeed, as if investigating this central problem of one of the most fully developed parts of any of the social sciences we have hit upon one of the essential differences between the social and the physical sciences. It is not the business of this essay to explore these more profound problems of methodology. But it may be suggested that if this case is at all typical – and some would regard the procedure of theory of prices as standing near the limit of proximity to the physical sciences – then the procedure of the social sciences which deal with conduct, which is in some sense purposive, can never be completely assimilated to the procedure of the physical sciences. It is really not possible to understand the concepts of choice, of the relationship of means and ends, the central concepts of our science, in terms of observation of external data. The conception of purposive conduct in this sense does not necessarily involve any ultimate indeterminism. But it does involve links in the chain of causal explanation which are psychical, not physical, and which are, for that reason, not necessarily susceptible of observation by behaviourist methods. Recognition of this does not in the least imply renunciation of "objectivity" in Max Weber's sense. It was exactly this that Max Weber had in mind when he wrote his celebrated essays.[25] All that the "objective" (that is to say, the *wertfrei*, to use Max Weber's phrase) explanation of conduct involves is the consideration of certain data, individual valuations, etc., which are not merely physical in character. The fact that such data are themselves of the nature of judgments of value does not necessitate that they should be valued as such. They are not judgments of value by the observer. What is of relevance to the social sciences is, not whether individual judgments of value are *correct* in the ultimate sense of the philosophy of value, but whether they are *made* and whether they are essential links in the chain of causal explanation. If the argument of this section is correct, this question must be answered in the affirmative. . . .

Chapter V: Economic Generalisations and Reality

. . .5. But to recognise that Economic laws are general in nature is not to deny the reality of the necessities they describe or to derogate from their value as a means of interpretation and prediction. On the contrary, having carefully delimited the nature and the scope of such generalisations, we may proceed with all the greater confidence to claim for them a complete necessity within this field.

Economic laws describe inevitable implications. If the data they postulate are given, then the consequences they predict necessarily follow. In this sense they are on the same footing as other scientific laws, and as little capable of "suspension". If, in a given situation, the facts are of a certain order, we are warranted in deducing with complete certainty that other facts which it enables us to describe are also present. To those who have grasped the implications of the propositions set forth in the last chapter the reason is not far to seek. If the "given situation" conforms to a certain pattern, certain other features must also be present, for their presence is "deducible" from the pattern originally postulated. The analytic method is simply a way of discovering the necessary consequences of complex collocations of facts – consequences whose counterpart in reality is not so immediately discernible as the counterpart of the original postulates. It is an instrument for "shaking out" all the implications of given suppositions. Granted the correspondence of its original assumptions and the facts, its conclusions are inevitable and inescapable.

All this becomes particularly clear if we consider the procedure of diagrammatic analysis. Suppose, for example, we wish to exhibit the effects on price of the imposition of a small tax. We make certain suppositions as regards the elasticity of demand, certain suppositions as regards the cost functions, embody these in the usual diagram, and we can at once *read off*, as it were, the effects on the price.[26] They are implied in the original suppositions. The diagram has simply made explicit the concealed implications.

It is this inevitability of economic analysis which gives it its very considerable prognostic value. It has been emphasised sufficiently already that Economic Science knows no way of predicting out of the blue the configuration of the data at any particular point of time. It cannot predict changes of valuations. But, given the data in a particular situation, it can draw inevitable conclusions as to their implications. And if the data remain unchanged, these implications will certainly be realised. They must be, for they are implied in the presence of the original data.

It is just here that we can perceive yet a further function for empirical investigation. It can bring to light the changing facts which make prediction in any given situation possible. As we have seen, it is most improbable that it can ever discover the law of their change, for the data are not subject to homogeneous causal influences. But it can put us in possession of information which is relevant at the particular moment concerned. It can give us some idea of the relative magnitude of the different forces operative. It can afford a basis for enlightened conjectures with regard to potential directions of change. And this unquestionably is one of the main uses of applied studies – not to unearth "empirical" laws in an area where such laws are not to be expected, but to provide from moment to moment some knowledge of the varying data on which, in the given situation, prediction can be based. It cannot supersede formal analysis. But it can suggest in different situations what formal analysis is appropriate, and it can provide *at that moment* some content for the formal categories.

Of course, if other things do not remain unchanged, the consequences predicted do not necessarily follow. This elementary platitude, necessarily implicit in *any* scientific prediction, needs especially to be kept in the foreground of attention when discussing this kind of prognosis. The statesman who said "*Ceteris paribus* be damned!" has a large and enthusiastic following among the critics of Economics! Nobody in his senses would hold that the laws of mechanics were invalidated if an experiment designed to illustrate them were interrupted by an earthquake. Yet a substantial majority of the lay public, and a good many *soi-disant* economists as well, are continually criticising well-established propositions on grounds hardly less slender.[27] A protective tariff is imposed on the importation of commodities, the conditions of whose domestic production make it certain that, if other things remain unchanged, the effect of such protection will be a rise in price. For quite adventitious reasons, the progress of technique, the lowering of the price of raw materials, wage reductions, of what not, costs are reduced and the price does not rise. In the eyes of the lay public and "Institutionalist" economists the generalisations of Economics are invalidated. The laws of supply and demand are suspended. The bogus claims of a science which does not regard the facts are laid bare. And so on and so forth. Yet, whoever asked of the practitioners of any other science that they should predict the complete course of an uncontrolled history?

Now, no doubt, the very fact that events in the large are uncontrolled,[28] that the fringe of given data is so extensive and so exposed to influence from unexpected quarters, must make the task of prediction, however carefully safeguarded, extremely hazardous. In many situations, small changes in particular groups of data are so liable to be counterbalanced by other changes which may be occurring independently and simultaneously, that the prognostic value of the knowledge of operative tendencies is small. But there are certain broad changes, usually involving many lines of expenditure or production at once, where a knowledge of implications is a very firm basis for conjectures of strong probability. This is particularly the case in the sphere of monetary phenomena. There can be no question that a quite elementary knowledge of the Quantity Theory was of immense prognostic value during the War and the disturbances which followed. If the speculators who bought German marks, after the War, in the confident expectation that the mark would automatically resume its old value, had been aware of as much of the theory of money as was known, say, to Sir William Petty, they would have known that what they were doing was ridiculous. Similarly, it becomes more and more clear, for purely analytical reasons, that, once the signs of a major boom in trade have made their appearance, the coming of slump and depression is almost certain; though when it will come and how long it will last are not matters which are predictable, since they depend upon human volitions occurring after the indications in question have appeared. So, too, in the sphere of the labour market, it is quite certain that some types of wage policy must result in unemployment if other things remain equal: and knowledge of how the "other things" must change in order that this consequence may be avoided makes it very often possible to predict with considerable confidence the actual results of given policies. These things have been verified again and again in practice. Today it is only he who is blind because he does not want to see who is prepared to deny them. If certain conditions are present, then, in the absence of new complications, certain consequences are inevitable. . . .

Chapter VI: The Significance of Economic Science

. . .2. It is sometimes thought that certain developments in modern Economic Theory furnish *by themselves* a set of norms capable of providing a basis for political practice. The Law of Diminishing Marginal Utility is held to provide a criterion of all forms of political and social activity affecting distribution. Anything conducive to greater equality, which does not adversely affect production, is said to be justified by this law; anything conducive to inequality, condemned. These propositions have received the support of very high authority. They are the basis of much that is written on the theory of public finance.[29] No less an authority than Professor Cannan has invoked them, to justify the ways of economists to Fabian Socialists.[30] They have received the widest countenance in numberless works on Applied Economics. It is safe to say that the great majority of English economists accept them as axiomatic. Yet with great diffidence I venture to suggest that they are in fact entirely unwarranted by any doctrine of scientific economics, and that outside this country they have very largely ceased to hold sway.

The argument by which these propositions are supported is familiar: but it is worth while repeating it explicitly in order to show the exact points at which it is defective. The Law of Diminishing Marginal Utility implies that the more one has of anything the less one values additional units thereof. Therefore, it is said, the more real income one has, the less one values additional units of income. Therefore the marginal utility of a rich man's income is less than the marginal utility of a poor man's income. Therefore, if transfers are made, and these transfers do not appreciably affect production, total utility will be increased. Therefore, such transfers are "economically justified". *Quod erat demonstrandum.*

At first sight the plausibility of the argument is overwhelming. But on closer inspection it is seen to be merely specious. It rests upon an extension of the conception of diminishing marginal utility into a field in which it is entirely illegitimate. The "Law of Diminishing Marginal Utility" here invoked does not follow in the least from the fundamental conception of economic goods; and it makes assumptions which, whether they are true or false, can never be verified by observation or introspection. The proposition we are examining begs the great metaphysical question of the scientific comparability of different individual experiences. This deserves further examination.

The Law of Diminishing Marginal Utility, as we have seen, is derived from the conception of a scarcity of means in relation to the ends which they serve. It assumes that, for each individual, goods can be ranged in order of their significance for conduct; and that, in the sense that it will be preferred, we can say that one use of a good is more important than another. Proceeding on this basis, we can compare the order in which one individual may be supposed to prefer certain alternatives with the order in which they are preferred by another individual. In this way it is possible to build up a complete theory of exchange.[31]

But it is one thing to assume that scales can be drawn up showing the *order* in which an individual will prefer a series of alternatives, and to compare the arrangement of one such individual scale with another. It is quite a different thing to assume that behind such arrangements lie magnitudes which themselves can be compared. This is not an assumption which need anywhere be made in modern economic analysis, and it is an assumption which is of an entirely different kind from the assumption of individual scales of relative valuation. The theory of exchange assumes that *I* can compare the importance *to me* of bread at 6d. per loaf and 6d. spent on other alternatives presented by the opportunities of the market. And it assumes that the order of my preferences thus exhibited can be compared with the order of preferences of the baker. But it does *not* assume that, at any point, it is necessary to compare the satisfaction which *I* get from the spending of 6d. on bread with the satisfaction which *the Baker* gets by receiving it. That comparison is a comparison of an entirely different nature. It is a comparison which is never needed in the theory of equilibrium and which is never implied by the assumptions of that theory. It is a comparison which necessarily falls outside the scope of any positive science. To state that A's preference stands above B's in order of importance is entirely different from stating that A prefers *n* to *m* and B prefers *n* and *m* in different order. It involves an element of conventional valuation. Hence it is essentially normative. It has no place in pure science.

If this is still obscure, the following consideration should be decisive. Suppose that a difference of opinion were to arise about A's preferences. Suppose that I thought that, at certain prices, he preferred *n* to *m*, and you thought that, at the same prices, he preferred *m* to *n*. It would be easy to settle our differences in a purely scientific manner. Either we could ask A to tell us. Or, if we refused to believe that introspection on A's part was possible, we could expose him to the stimuli in question and observe his behaviour. Either test would be such as to provide the basis for a settlement of the difference of opinion.

But suppose that we differed about the satisfaction derived by A from an income of £1,000, and the satisfaction derived by B from an income of twice that magnitude. Asking them would provide no solution. Supposing they differed. A might urge that he had more satisfaction than B at the margin. While B might urge that, on the contrary, he had more satisfaction than A. We do not need to be slavish behaviourists to realise that here is no scientific evidence. *There is no means of testing the magnitude of A's satisfaction as compared with B's.* If we tested the state of their blood-streams, that would be a test of blood, not satisfaction. Introspection does not enable A to measure what is going on in B's mind, nor B to measure what is going on in A's. There is no way of comparing the satisfactions of different people.

Now, of course, in daily life we do continually assume that the comparison can be made. But the very diversity of the assumptions actually made at different times and in different places is evidence of their conventional nature. In Western democracies we assume for certain purposes that men in similar circumstances are capable of equal satisfactions. Just as for purposes of justice we assume equality of responsibility in similar situations as between legal subjects, so for purposes of public finance we agree to assume equality of capacity for experiencing satisfaction from equal incomes in similar circumstances as between economic subjects. But, although it may be convenient to assume this, there is no way of proving that the assumption rests on ascertainable fact. And, indeed, if the representative of some other civilisation were to assure us that we were wrong, that members of his caste (or his race) were capable of experiencing ten times as much satisfaction from given incomes as members of an inferior caste (or an "inferior" race), we could not refute him. We might poke fun at him. We might flare up with indignation, and say that his valuation was hateful, that it led to civil strife, unhappiness, unjust privilege, and so on and so forth. But we could not show that he was wrong in any objective sense, any more than we could show that we were right. And since in our hearts we do not regard different men's satisfactions from similar means as equally valuable, it would really be rather silly if we continued to pretend that the justification for our scheme of things was in any way *scientific.* It can be justified on grounds of general convenience. Or it can be justified by appeal to ultimate standards of obligation. But it cannot be justified by appeal to any kind of positive science.

Hence the extension of the Law of Diminishing Marginal Utility, postulated in the propositions we are examining, is illegitimate. And the arguments based upon it therefore are lacking in scientific foundation Recognition of this no doubt involves a substantial curtailment of the claims of much of what now assumes the status of scientific generalisation in current discussions of Applied Economics. The conception of diminishing relative utility (the convexity downwards of the indifference curve) does not justify the inference that transferences from the rich to the poor will increase total satisfaction. It does not tell us that a graduated income tax is less injurious to the social dividend than a nongraduated poll tax. Indeed, all that part of the theory of public finance which deals with "Social Utility" must assume a different significance. Interesting as a development of an ethical postulate, it does not at all follow from the positive assumptions of pure theory. It is simply the accidental deposit of the historical association of English Economics with Utilitarianism: and both the utilitarian postulates from which it derives and the analytical Economics with which it has been associated will be the better and the more convincing if this is clearly recognised.[32]

But supposing this were not so. Suppose that we could bring ourselves to believe in the positive status of these conventional assumptions, the commensurability of different experiences, the equality of capacity for satisfaction, etc. And suppose that, proceeding on this basis, we had succeeded in showing that certain policies *had the effect* of increasing "social utility", even so it would be totally illegitimate to argue that such a conclusion by itself warranted the inference that these policies *ought* to be carried out. For such an inference would beg the whole question whether the increase of satisfaction in this sense was socially obligatory.[33] And there is nothing within the body of economic generalisations, even thus enlarged by the inclusion of elements of conventional valuation, which affords any means of deciding this question. Propositions involving "ought" are on an entirely different plane from propositions involving "is". . . .

5. But what, then, is the significance of Economic Science? We have seen that it provides, within its own structure of generalisations, no norms which are binding in practice. It is incapable of deciding as between the desirability of different ends. It is fundamentally distinct from Ethics. Wherein, then, does its unquestionable significance consist?

Surely it consists in just this, that, when we are faced with a choice between ultimates, it enables us to choose with full awareness of the implications of what we are choosing. Faced with the problem of deciding between this and that, we are not entitled to look to Economics for the ultimate decision. There is nothing in Economics which relieves *us* of the obligation to choose. There is nothing in any kind of science which can decide the ultimate problem of preference. But, to be completely rational, we must know what it is we prefer. We must be aware of the implications of the alternatives. For rationality in choice is nothing more and nothing less than choice with complete awareness of the alternatives rejected. And it is just here that Economics acquires its practical significance. It can make clear to us the implications of the different ends we may choose. It makes it possible for us to will with knowledge of what it is we are willing. It makes it possible for us to select a system of ends which are mutually consistent with each other.[34]

An example or two should make this quite clear. Let us start with a case in which the implications of one act of choice are elucidated. We may revert once more to an example we have already considered – the imposition of a protective tariff. We have seen already that there is nothing in scientific Economics which warrants our describing such a policy as good or bad. We have decided that, if such a policy is decided upon with full consciousness of the sacrifices involved, there is no justification for describing it as uneconomical. The deliberate choice by a body of citizens acting collectively to frustrate, in the interests of ends such as defence, the preservation of the countryside, and so on, their several choices as consumers, cannot be described as uneconomical or irrational, if it is done with full awareness of what is being done. But this will not be the case unless the citizens in question are fully conscious of the objective implications of the step they are taking. And in an extensive modern society it is only as a result of intricate economic analysis that they may be placed in possession of this knowledge. The great majority, even of educated people, called upon to decide upon the desirability of, let us say, protection for agriculture, think only of the effects of such measures on the protected industry. They see that such measures are likely to benefit the industry, and hence they argue that the measures are good. But, of course, as every first year student knows, it is only here that the problem begins. To judge the further repercussions of the tariff an analytical technique is necessary. This is why in countries where the level of education in Economics is not high, there is a constant tendency to the approval of more and more protective tariffs.

Nor is the utility of such analysis to be regarded as confined to decisions on isolated measures such as the imposition of a single tariff. It enables us to judge more complicated systems of policy. It enables us to see what *sets* of ends are compatible with each other and what are not, and upon what conditions such compatibility is dependent. And, indeed, it is just here that the possession of some such technique becomes quite indispensable if policy is to be rational. It may be just possible to will rationally the achievement of particular social ends overriding individual valuations without much assistance from analysis. The case of a subsidy to protect essential food supplies is a case in point. It is almost impossible to conceive the carrying through of more elaborate policies without the aid of such an instrument.[35]

We may take an example from the sphere of monetary policy. It is an unescapable deduction from the first principles of monetary theory that, in a world in which conditions are changing at different rates in different monetary areas, it is impossible to achieve at once stable prices and stable exchanges.[36] The two ends – in this case the "ends" are quite obviously subordinate to other major norms of policy – are logically incompatible. You may try for one or you may try for the other – it is not certain that price stability is either permanently attainable or conducive to equilibrium generally – but you cannot rationally try for both. If you do, there must be a breakdown. These conclusions are well known to all economists. Yet without some analytical apparatus how few of us would perceive the incompatibility of the ends in question!

And even this is a narrow example. Without economic analysis it is not possible rationally to choose between alternative *systems* of society. We have seen already that if we regard a society which permits inequality of incomes as an evil in itself, and an equalitarian society as presenting an end to be pursued above all other things, then it is illegitimate to regard such a preference as uneconomic. But it is not possible to regard it as rational unless it is formulated with a full consciousness of the nature of the sacrifice which is thereby involved. And we cannot do this unless we understand, not only the essential nature of the capitalistic mechanism, but also the necessary conditions and limitations to which the type of society proposed as a substitute would be subject. It is not rational to will a certain end if one is not conscious of what sacrifice the achievement of that end involves. And, in this supreme weighing of alternatives, only a complete awareness of the implications of modern economic analysis can confer the capacity to judge rationally.

But, if this is so, what need is there to claim any larger status for Economic Science? Is it not the burden of our time that we do not realise what we are doing? Are not most of our difficulties due to just this fact, that we will ends which are incompatible, not because we wish for deadlock, but because we do not realise their incompatibility. It may well be that there may exist differences as regards ultimate ends in modern society which render some conflict inevitable. But it is clear that many of our most pressing difficulties arise, not for this reason, but because our aims are not co-ordinated. As consumers we will cheapness, as producers we choose security. We value one distribution of factors of production as private spenders and savers. As public citizens we sanction arrangements which frustrate the achievement of this distribution. We call for cheap money and lower prices, fewer imports and a larger volume of trade.[37] The different "will-organisations" in society, although composed of the same individuals, formulate different preferences. Everywhere our difficulties seem to arise, not so much from divisions between the different members of the body politic, as from, as it were, split personalities on the part of each one of them.[38]

To such a situation, Economics brings the solvent of knowledge. It enables us to conceive the far-reaching implications of alternative possibilities of policy. It does not, and it cannot, enable us to evade the necessity of choosing between alternatives. But it does make it possible for us to bring our different choices into harmony. It cannot remove the ultimate limitations on human action. But it does make it possible within these limitations to act consistently. It serves for the inhabitant of the modern world with its endless interconnections and relationships as an extension of his perceptive apparatus. It provides a technique of rational action.

This, then, is a further sense in which Economics can be truly said to assume rationality in human society. It makes no pretence, as has been alleged so often, that action is necessarily rational in the sense that the ends pursued are not mutually inconsistent. There is nothing in its generalisations which necessarily implies reflective deliberation in ultimate valuation. It relies upon no assumption that individuals will always act rationally. But it does depend for its practical *raison d'être* upon the assumption that it is desirable that they should do so. It does assume that, within the bounds of necessity, it is desirable to choose ends which can be achieved harmoniously.

And thus in the last analysis Economics does depend, if not for its existence, at least for its significance, on an ultimate valuation – the affirmation that rationality and ability to choose with knowledge is desirable. If irrationality, if the surrender to the blind force of external stimuli and uncoordinated impulse at every moment is a good to be preferred above all others, then it is true the *raison d'être* of Economics disappears. And it is the tragedy of our generation, red with fratricidal strife and betrayed almost beyond belief by those who should have been its intellectual leaders, that there have arisen those who would uphold this ultimate negation, this escape from the tragic necessities of choice which has become conscious. With all such there can be no argument. The revolt against reason is essentially a revolt against life itself. But for all those who still affirm more positive values, that branch of knowledge which, above all others, is the symbol and safeguard of rationality in social arrangements, must, in the anxious days which are to come, by very reason of this menace to that for which it stands, possess a peculiar and a heightened significance.

Notes

1. It is perhaps worth emphasising the significance of this qualification. The application of technically similar means to the achievement of qualitatively similar ends at *different times* constitutes alternative uses of these means. Unless this is clearly realised, one of the most important types of economic action is overlooked.
2. This may seem an unnecessary refinement, and in the first edition of this essay I left it out for that reason. But the condition that there exists a hierarchy of ends is so important in the theory of value that it seems better to state it explicitly even at this stage. See Chapter IV., Section 2.
3. Cp. Schönfeld, *Grenznutzen und Wirtschaftsrechnung*, p. 1; Hans Mayer, *Untersuchungen zu dem Grundgesetze der wirtschaftlichen Wertrechnung* (*Zeitschrift für Volkswirtschaft und Sozialpolitik*, Bd. 2, p. 123).

   It should be sufficiently clear that it is not "time" as such which is scarce, but rather the potentialities of ourselves viewed as instruments. To speak of scarcity of time is simply a metaphorical way of invoking this rather abstract concept.
4. It should be clear that there is no disharmony between the conception of end here employed, the terminus of particular lines of conduct in acts of final consumption, and the conception involved when it is said that there is but one end of activity – the maximising of satisfaction, "utility", or what not. Our "ends" are to be regarded as proximate to the achievement of this ultimate end. If the means are scarce they cannot all be achieved, and according to the scarcity of means and their relative importance the achievement of some ends has to be relinquished.
5. Menger, *Grundsütze der Volkswirtschaftslehre*, lte Aufl., pp. 51–70; Mises, *Die Gemeinwirtschaft*, pp. 98 *seq.*; Fetter, *Economic Principles*, ch. i.; Strigl, *Die ökonomischen Kategorien und die Organisation der Wirtschaft, passim*; Mayer, op. cit.
6. See below, Section 4, and Chapter V., Section 3.
7. For an example of such a derivation reaching substantially similar results, see Strigl, op. cit., pp. 121 seq.
8. If any reader of this book has any doubt of the evidence of the facts, he should consult the standard work on recent British experiments in such measures, *British Food Control*, by Sir William Beveridge.
9. "The vulgar notion that the safe methods on political subjects are those of Baconian induction – that the true guide is not general reasoning but specific experience – will one day be quoted as among the most unequivocal marks of a low state of the speculative faculties of any age in which it is accredited. . . . Whoever makes use of an argument of this kind . . . should be sent back to learn the elements of some one of the more easy physical sciences. Such reasoners ignore the fact of Plurality of Causes in the very case which affords the most signal example of it" (John Stuart Mill, *Logic*, chapter x., paragraph 8).

10. See Rickert, op. cit., pp. 78–101, *Die Grenzen der naturwissenschaftlichen Begriffsbildung, passim.*
11. On all this see the illuminating observations of Dr. Strigl, *Die ökonomischen Kategorien und die Organisation der Wirtschaft*, pp. 85–121.
12. See Robinson, *Economics of Imperfect Competition*, pp. 330–1. I myself first learnt this way of putting things from a conversation with Professor Mises many years ago. But so far as I know Mrs. Robinson is the first to put matters so succinctly and clearly in print: I think that Mrs. Robinson's book will have done much to convince many hitherto sceptics of the utility and significance of the kind of abstract reasoning from very simple postulates which is the subject of the present discussion.
13. See, *e.g.*, Schneider, *Theorie des Produktion, passim.*
14. See Knight, *Risk, Uncertainty, and Profit*; Hicks, *The Theory of Profit* (*Economica*, No. 31, pp. 170–90).
15. See Mises, *The Theory of Money*, pp. 147 and 200; Lavington, *The English Capital Market*, pp. 29–35; Hicks, *A Suggestion for Simplifying the Theory of Money* (*Economica*, 1934, pp. 1–20).
16. Cp. Mises, *Kritik des Interventionismus*, pp. 55–90.
17. *Economic Principles*, pp. ix and 12–21.
18. See Menger, *Grundsätze*, 1 Aufl., pp. 77–152.
19. See *Positive Theorie des Kapitals*, 4[e] Auflage, pp. 232–46.
20. *Mathematical Psychics*, p. 15.
21. For a refutation of this view, see Max Weber, *Die Grenznutzenlehre und das psychophysische Grundgesetz (Archiv für Sozialwissenschaft und Sozialpolitik*, vol. xxix., 1909).
22. *The Theory of Social Economy*, First English Edition, vol. i., pp. 50–1.
23. Notably in the article on *Economie mathématique* in the *Encyclopédie des Sciences mathématiques*, Paris, 1911.
24. See, *e.g.*, Hicks, *Gleichgewicht und Konjunktur* (*Zeitschrift für Nationalökonomie*, vol. iv., pp. 441–55).
25. Max Weber, *Die Objectivität socialwissenschaftlichen und socialpolitischen Erkenntnis: Der Sinn der Wertfreiheit der soziologischen und ökonomischen Wissenschaft* in *Gesammelte Aufsatze zur Wissenschaftlehre.*
26. See, *e.g.*, Dalton, *Public Finance*, 2nd edition, p. 73.
27. See, *e.g.*, the various statistical "refutations" of the quantity theory of money which have appeared in recent years. On all these the comment of Torrens on Tooke is all that need be said. "The History of Prices may be regarded as a psychological study. Mr. Tooke commenced his labours as a follower of Horner and Ricardo, and derived reflected lustre from an alliance with those celebrated names; but his capacity for collecting contemporaneous facts preponderating over his perceptive and logical faculties, his accumulation of facts involved him in a labyrinth of error. Failing to perceive that a theoretical principle, although it may irresistibly command assent under all circumstances coinciding with the premises from which it is deduced, must be applied with due limitation and correction in all cases not coinciding with the premises, he fell into a total misconception of the proposition advanced by Adam Smith, and imputed to that high authority the absurdity of maintaining that variations in the quantity of

money cause the money values of all commodities to vary in equal proportions, while the values of commodities, in relation to each other, are varying in unequal proportions. Reasonings derived from this extraordinary misconception necessarily led to extraordinary conclusions. Having satisfied himself that Adam Smith had correctly established as a principle universally true that variations in the purchasing power of money cause the prices of all commodities to vary in equal proportions, and finding, as he pursued his investigations into the phenomena of the market at different periods, no instances in which an expansion or contraction of the circulation caused the prices of commodities to rise or fall in an equal ratio, he arrived by a strictly logical inference from the premises thus illogically assumed, at his grand discovery – that no increase of the circulating medium can have the effect of increasing prices" (*The Principles and Operation of Sir Robert Peel's Act of 1844 Explained and Defended*, 1st edition, p. 75).

28. The alleged advantage of economic "planning" – namely, that it enables greater certainty with regard to the future – depends upon the assumption that under "planning" the present controlling forces, the choices of individual spenders and savers, are themselves brought under the control of the planners. The paradox therefore arises that either the planner is destitute of the instrument of calculating the ends of the community he intends to serve, or, if he restores the instrument, he removes the *raison d'être* of the "plan". Of course, the dilemma does not arise if he thinks himself capable of interpreting those ends or – what is much more probable – if he has no intention of serving any other ends but those *he* thinks appropriate. Strange to say this not infrequently happens. Scratch a would-be planner and you usually find a would-be dictator.
29. See, *e.g.*, Edgeworth, *The Pure Theory of Taxation* (*Papers Relating to Political Economy*, vol. ii., pp. 63 *seq.*).
30. See *Economics and Socialism* (*The Economic Outlook*, pp. 59–62).
31. So many have been the misconceptions based upon an imperfect understanding of this generalisation that Dr. Hicks has suggested that its present name be discarded altogether and the title Law of Increasing Rate of Substitution be adopted in its place. Personally, I prefer the established terminology, but it is clear that there is much to be said for the suggestion.
32. Cp. Davenport, *Value and Distribution*, pp. 301 and 571; Benham, *Economic Welfare* (*Economica*, June, 1930, pp. 173–87); M. St. Braun, *Theorie der staatlichen Wirtschaftspolitik*, pp. 41–4. Even Professor Irving Fisher, anxious to provide a justification for his statistical method for measuring "marginal utility", can find no better apology for his procedure than that "Philosophic doubt is right and proper, but the problems of life cannot and do not wait" (*Economic Essays in Honour of John Bates Clark*, p. 180). It does not seem to me that the problem of measuring marginal utility as between individuals is a particular pressing problem. But whether this is so or not, the fact remains that Professor Fisher solves his problem only by making a conventional assumption. And it does not seem that it anywhere aids the solution of practical problems to pretend that conventional assumptions have scientific justification. It does not make me a more docile democrat to be told that *I* am equally capable of experiencing satisfaction as my neighbour; it fills me with indignation. But I am perfectly willing to accept the statement that it is *convenient* to assume that this is the case.

I am quite willing to accept the argument – indeed, as distinct from believers in the racial or proletarian myths, I very firmly believe – that, in modern conditions, societies which proceed on any other assumption have an inherent instability. But we are past the days when democracy could be made acceptable by the pretence that judgments of value are judgments of scientific fact. I am afraid that the same strictures apply to the highly ingenious *Methods for Measuring Marginal Utility* of Professor Ragnar Frisch.

33. Psychological hedonism in so far as it went beyond the individual may have involved a non-scientific assumption, but it was not by itself a necessary justification for ethical hedonism.
34. It is perhaps desirable to emphasise that the consistency which is made possible is a consistency of achievement, not a consistency of ends. The achievement of one end may be held to be inconsistent with the achievement of another, either on the plane of, valuation, or on the plane of objective possibility. Thus it may be held to be ethically inconsistent to serve two masters at once. It is objectively inconsistent to arrange to be with each of them at the same time, at different places. It is the latter kind of inconsistency in the sphere of social policy which scientific Economics should make it possible to eliminate.
35. All this should be a sufficient answer to those who continually lay it down that "social life is too complex a matter to be judged by economic analysis". It is because social life is so complicated that economic analysis is necessary if we are to understand even a part of it. It is usually those who talk most about the complexity of life and the insusceptibility of human behaviour to any kind of logical analysis who prove to have the most *simpliste* intellectual and emotional make-up. He who has really glimpsed the irrational in the springs of human action will have no "fear" that it can ever be killed by logic.
36. See Keynes, *A Tract on Monetary Reform*, pp. 154–5; also an interesting paper by Mr. D. H. Robertson, *How do We Want Gold to Behave?* reprinted in the *International Gold Problem*, pp. 18–46.
37. Cf. M. S. Braun, *Theorie der Staatlichen Wirtschaftspolitik*, p. 5.
38. In this way economic analysis reveals still further examples of a phenomenon to which attention has often been drawn in recent discussion of the theory of Sovereignty in Public Law. See Figgis, *Churches in the Modern State*; Maitland, *Introduction to Gierke's Political Theories of the Middle Ages*; Laski, *The Problem of Sovereignty, Authority in the Modern State.*

FOUR

Economics and Human Action

Frank Knight

Frank Knight (1885–1972) was born in Atlanta and received his Ph.D. in economics from Cornell. He taught at Cornell and at the University of Iowa, but he is particularly associated with the University of Chicago, where he trained a whole generation of prominent economists. Knight not only made major contributions to economic theory, but he was a social philosopher as well, deeply concerned with the problems of individual liberty. His insistence on the importance of uncertainty and on the peculiarities of the human subject matter of economics is still worth careful consideration by all those interested in economic methodology. Knight's essay, "Value and Price," is reprinted here in an abridged form. The first third of the essay, which is mostly historical background, is omitted.

In general, if explanation of economic behaviour in terms of motives is to be abandoned, a number of alternative possibilities are open. Perhaps the simplest is the one analogous to a trend in physics – to do away with all "explanation" and merely to formulate empirical laws; the result is statistical economic theory, having for its content the objective phenomena of commodities and prices alone. A second line of development away from the types of value theory represented by classical or utility economics centres around the emphasis on the social control of economic life with clearly implied advocacy of such control. In the past generation this trend has been most marked in Germany (socialism of the chair), in England (Fabianism and left wing liberalism), and in the United States (as a phase of institutionalist economics).

The third alternative to explanatory theory is that of treating economic phenomena as essentially historical, which, of course, must be done in any case if the concrete content of economic life at a particular time and place is to be explained. Historical economics again subdivides into as many varieties as

---

From *The Ethics of Competition and Other Essays*, by Frank Knight. New York and London: Harper and Brothers, 1935.

there are basic conceptions of history and historical method. Two such varieties stand out. The first treats history as far as possible in objective, empirical terms, and may use statistics for the discovery and analysis of trends; logically this procedure contrasts sharply with the search for repetitive laws, analogous to those of natural science, which characterizes statistical economic theory, but in practice the two conceptions run together in the work of statistical economists. The second variety of historical economics uses the more familiar humanistic conceptions of political and social history – individual ambitions, efforts, and failures in a given social-psychological setting. It represents essentially a revival or continuation of the historical schools of the nineteenth century, especially prominent in Germany. In so far as it arrives at generalization, it may be described as institutional economics, a term which has come into use particularly in the United States. The related contemporary movement in the German literature is referred to as neohistorical or sociological economics, with Sombart and Max Weber as its most prominent leaders.

At the root of the differences and disputes between the old and the new economics as well as among the three new lines of theoretical development noted above are two problems: the relation between description and explanation and the relation between statement of fact and critical evaluation. The first, inescapable in any thinking about human conduct, is fundamentally the problem of the reality of choice, or "freedom of the will." It involves the essence of the value problem in the sense of individual values, and is at bottom the problem of the relation between individual man and nature. The second basic problem has to do with the relation between the individual man and society.

The crucial fact in connection with the first problem is that, if motive or end in any form is granted any real role in conduct, it cannot be that of a cause in the sense of causality in natural science. This is the supreme limitation alike of statistical and historical economics. For, if motive or end is used to explain behaviour, it must in turn be brought into the same relation with events and conditions antecedent to it, and then the motive becomes superfluous; the behaviour will be fully accounted for by these antecedents. Motive cannot be treated as a natural event. A fundamental contrast between cause and effect in nature and end and means in human behaviour is of the essence of the facts which set the problem of interpreting behaviour. There seems to be no possibility of making human problems real, without seeing in human activity an element of effort, contingency, and, most crucially, of error, which must for the same reasons be assumed to be absent from natural processes.

Thus motive or intent forces itself into any relevant discussion of human activity. But the subject of behaviour cannot be simplified even to the point of reducing it to a dualism. At least three basic principles must be introduced into its interpretation. The typical human action is explained in part by natural causality, in part by an intention or desire which is an absolute datum and is thus a "fact" although not a natural event or condition, and in part by an urge to realize "values" which cannot be reduced entirely to factual desires because this urge has no literally describable objects. Interpretation in terms of factual desires is the procedure of economics as represented by the bulk of the theoretical literature, in so far as it is objective in outlook. Yet this second principle of explanation is perhaps the most vulnerable of the three. It is doubtful whether any desire is really "absolute," whether there exists any desire that does not look to achievement of some change in a growing system of meaning and values; this is a different thing from changes in physical nature, even though rearrangements in physical nature are the only means by which values can be realized. Every act, in the economic sense, changes the configuration of matter in space. But this does not exclude the possibility of "acts" which change meaning and values without changing natural configuration, since reflection may yield new insight and effect a change of personal tastes. More fundamentally, it is doubtful whether one configuration is in itself preferable to another.

People report and feel two different types of motivation for their acts. There is the wish or preference which is treated by the actor and by outsiders as final, as a brute fact. On the other hand, people make value judgments of various sorts in explanation of their acts; and explanation runs into justification. In other words, no one can really treat motive objectively or describe a motive without implications of good and bad. Thus not only do men desire more or less distinctly from valuing, but they desire because they value and also value without desiring. Indeed, the bulk of human valuations, in connection with truth, beauty, and morals, are largely or altogether independent of desire for any concrete thing or result. That individual economic motivation itself typically involves some valuation and not merely desire is established by two other considerations: first, what is chosen in an economic transaction is generally wanted as a means to something else, which involves a judgment that it "really" is a means to the result in question; and, second, what is ultimately wanted for its own sake can rarely, if ever, finally be described in terms of physical configuration, but must be defined in relation to a universe of meanings and values. Thus there is an element of valuation in the notion of efficiency in the realization of a given end; and, in addition, the real end contains as an element a value concept.

The dual conception found in motivation is reflected also in the more narrowly economic concept of value. The latter contains definitely more than the notion of a quality measured by price; it is always imperfectly measured under actual conditions. Price "tends" to coincide with value, but the notion of value also involves a norm to which price would conform under some ideal conditions. This norm includes two ideas: that of a goal aimed at but only more or less approximately realized because of errors of various kinds (which tend to be corrected); and that of a "correct" goal of action in contrast with incorrect goals as well as the actual goal. In a society based upon competition as an accepted principle, the competitive price, or price equal to necessary costs of production, is the true value in both senses; aberrations are to be attributed to two sets of cases – accidental miscalculations, and wrong objectives of action. This statement overlooks, of course, the existence of different technical conceptions of competitive price relative to the short run or local conditions; and a deeper ethical criticism may condemn given conditions other than the tastes of consumers which fix competitive price, especially the distribution of income and economic power.

To make the main point clear it is necessary to notice the difference in the conception of ideal conditions in economics and in mechanics. In the latter field the most notable of the ideal conditions is the absence of friction; an apparently similar conception of ideal conditions is one of the familiar features, almost a cliche, in economic theory. As generalized description the conception of perfect competition, reached by abstraction from the features of the economic situation which make competition imperfect, is like the conceptions of frictionless mechanics and is similarly justified. But to assume that the specific thing abstracted from in the theory of perfect competition bears the same relation to behaviour as does friction to mechanical process would be utterly misleading. Friction in mechanics involves a transformation of energy from one form to another, according to a law just as rigid and a conservation principle just as definite as the law and conservation principle which hold good for mechanical changes where no energy disappears. There is nothing corresponding to any of this in the economic process. What is abstracted in equilibrium price theory is the fact of error in economic behaviour. Perfect competition is, among other things irrelevant here, errorless competition; fundamentally it is not comparable to a frictionless machine. The familiar "tendency" of competition to conform to the theoretical ideal is no mere possibility of experimental approximation, but a real tendency in so far as men are supposed to endeavour with some success to learn to behave intelligently. It cannot be treated as a tendency toward an objective result, but only as a tendency to conformity with the intent of behaviour, which intent cannot be measured or identified or defined in terms of any experimental data. The ideal conditions of economics involve perfect valuation in a limited sense, perfect economic behaviour which assumes the end or intention as given. The correctness of the intention is an ethical question, from which the economist abstracts just as he abstracts from error which causes the behaviour to end otherwise than according to the intent.

Thus far two levels of interpretation of economic behaviour have been discussed. The first is that at which behaviour is reduced as far as possible to principles of regularity by statistical procedure; it may or may not be thought convenient to impute behaviour to some "force," but if it is so adjudged, the force must be assumed to correspond with the behaviour observed. The second is the interpretation of behaviour in terms of motivation, which must centre on the difference between motive and act and on the fact of error. It is at the third level of interpretation that the intentional end of action itself is submitted to valuation or criticism from some point of view. Here the relation between individual and society, the second main problem suggested above, and the concept of value as related to social policy become central topics of discussion.

In fact even at the second level two forms of social reference must be recognized: the individual ends as they are given are chiefly social in origin and content; and in societies in which economic thinking has any relevance there is a large social-ethical acceptance and approval of individual motivation in the abstract. Modern society, for instance, has accepted the right and even the duty of the individual to pursue his own ends within wide limits; in other words, individual liberty itself is a social value and not merely a fact. Thus the second level of interpretation tends to break down. If the notion of economic behaviour is effectively separated from mechanical process, if the ends are regarded as ends and not merely as physical effects, the discussion is already in large part at the third level. Factual ends as desired cannot be maintained unless they are given a large element of valuation in addition to desire. The "desires" for economic goods and services cannot be held to be final or to have a self-contained, independent reality. The least scrutiny shows that they are very largely rather accidental manifestations of desire for something of the nature of liberty or power. But such objects of desire are forms of social relationship and not things, and the notion of economic efficiency has only a limited applicability to their pursuit and attainment. Treatment of such activities, if it is to have any general, serious appeal, must be a discussion of social policy relative to social ends or norms and social procedure in realizing them.

The serious difficulty in economic theory in this connection has been the tendency to confuse advocacy of a policy of political noninterference (or the opposite) with description of a social organization based on free contract. Even when the authors have not deliberately intended to preach as well as to analyse, the difficulties of keeping the two types of discussion separate have been too great, especially in view of the requirements of an exposition which would be intelligible, not to say appealing, to any considerable reading public. In this field the interest in values, and especially in social policy, is in fact predominant. Thus economic theory, growing up in an atmosphere of reaction against control, clearly overemphasized this side of the case and neglected the other. It is now just as obvious that there are equally sweeping and complex limitations to the principle of liberty in the economic sense, that is, to the organization of economic life exclusively through free contract among individuals using given resources to achieve given individual ends. Society cannot accept individual ends and individual means as data or as the main objectives of its own policy. In the first place, they simply are not data, but are historically created in the social process itself and are inevitably affected by social policy. Secondly, society cannot be even relatively indifferent to the workings of the process. To do so would be ultimately destructive of society and individual alike. This conclusion is strongly reinforced by the fact that the immediate interest of the individual is largely competitive, centered in his own social advancement relative to other individuals. In such a contest it is the function of the public authority to enforce the rules impartially, and still more to make such rules as would tend to keep the "game" on the highest possible level. To this end it must maintain a standpoint distinctly different from the interest in which the individual, always more conscious of conflicts of interests than of community of interest with the social body as a whole, tends to be absorbed.

These reflections point to a logical error underlying the value theory typical of the classical economists. It was not ostensibly their contention that liberty as such is a good. Notoriously, they were hedonists; their argument for liberty made it instrumental to pleasure, on the ground that the individual is a better judge than government officials of the means to his happiness. It is not denying weight to this argument to point out that liberty itself is unquestionably a good to the individual, and in addition an ethical good more or less apart from the degree to which the individual actually prizes it. Certainly an individual may desire liberty and claim a right to it without contending that he will uniformly make decisions more wisely than they would be made for him, from the standpoint of his own material comfort and security. And just as certainly it can be maintained that the individual should within limits make his own decisions and abide by their consequences even if he may not choose to do so. In other words, the classical economists did not realize, and the "scientific" spirit of the age has made economists generally reluctant to admit that liberty is essentially a social value, at least when it is advocated or opposed, as is any other social system or social relation.

The actual interests or desires expressed in economic behaviour are to an overwhelming extent social in genesis and in content; consequently they cannot be described apart from a system of social relations which itself cannot be treated in purely objective, factual terms. To a limited extent they can be conceived by an individual in such terms; they may even be described by one individual to another as matter of fact. But the parties to such a communication place themselves in the role of spectators rather than members of society or participants in the phenomena. Thus any published discussion, presupposing a general appeal to readers as members and participants, necessarily takes the form of stating a case for a policy, possibly with more or less equal attention to both sides. In this conflict between the spectator's interest in seeing and understanding and the participant's interest in action and change, the philosopher or methodologist cannot possibly take sides. The question whether economics as such should be one or the other is to be answered only by recognition that it must be both, with more or less emphasis one way or the other according to the aims of a particular treatment; but always by implication it must be both, however one-sided the emphasis, since each interest presupposes and is relative to the other, and every writer and reader as a human being is motivated by both interests. What is desirable is that in any statement the relation between the two sets of interests should be clear. But what tends to happen is the reverse: he whose interest is primarily in truth tends to reinforce his statements by identifying truth and value, and he whose interest is in values tends to strengthen his statements by giving them the quality of truth.

While in the period of development of the classical economics the practical social interest centred almost exclusively on liberation from an antiquated system of control, at present the pendulum has swung definitely the other way. The new problem raised by the confusion of scientific and evaluative interests is enormously more difficult than the old. Society is positively seeking a basis of unity and order instead of negatively attempting to abandon an unsatisfactory basis. Moreover, the current standards of thinking have come under extreme domination of the scientific ideal, which has little if any applicability to the problem. The ultimate foundation of group unity must be of the nature of morale and sentiment rather than knowledge. There is no intellectual solution of conflicts of interest. Only values can be discussed, but the discussion does not necessarily lead to agreement; and disagreement on principles seems morally to call for an appeal to force. It is also of interest to note that the tendency to "rationalization" causes conflict of interest and disagreement regarding principles each to take on the quality of its opposite, and that in practice they are inseparably mingled.

The extremist wings in the advocacy of change recognize the inapplicability of purely intellectual knowledge. Both "fascist" and "communist" schools incline to treat the truth or falsity of propositions in economics as a matter of indifference or even as illusory, judging the doctrines only by their conduciveness toward the establishment of the desired type of social order. This view is, of course, "untrue" from a narrower "scientific" point of view; in any social order the results of certain choices affecting production and consumption, by whomever made, come under certain abstract, essentially mathematical principles which express the difference between economy and waste. At the other extreme – at the first and second levels of interpretation indicated above – there is an equally energetic movement in the interest of a rigorously "scientific" treatment of economics. Analysis at the first level, disregarding motivation and considering only the results of action in the form of commodity statistics, leaves no real place for any concept of economy. Moreover, it cannot be carried out even literally, for commodities must be named and classified and the treatment must take account of similarities and differences in use as well as physical characteristics. And economics at the second level, treating desires as facts, is subject to very narrow limitations. Desires really have no very definite content, and of what they have the students can have no definite knowledge. The conception can be made the basis of a purely abstract theory, but it has little application to reality. To give the data any content, the desires must be identified with the goods and services in which they find expression, and the second method then is reduced to identity with the first. Moreover, the only desires which can be treated as at all akin to scientific data are purely individual, and any discussion of social policy must draw on values or ideals entirely outside of such a system.

FIVE

Selected Texts on Economics, History, and Social Science

Karl Marx

Reprinted here are three texts. The first, "Estranged Labour" from Marx's *Economic and Philosophical Manuscripts of 1844*, provides a sweeping overview of his vision of the way in which the economic relations among people and the products of those relations dominate the very people who create and sustain those relations. The second, Marx's "Preface" to *A Contribution to the Critique of Political Economy*, very briefly sketches Marx's historical materialism, whereby the state of technology determines the economic relations among people, which in turn determine legal and political relations and the course of history. The third, "The Method of Political Economy," which is a section of the "Introduction" to *A Contribution to the Critique of Political Economy*, contains Marx's most explicit and sustained discussion of economic methodology.

Estranged Labour

We have started out from the premises of political economy. We have accepted its language and its laws. We presupposed private property; the separation of labour, capital, and land, and likewise of wages, profit, and capital; the division of labour; competition; the conception of exchange value, etc. From political economy itself, using its own words, we have shown that the worker sinks to the level of a commodity, and moreover the most wretched commodity of all; that the misery of the worker is in inverse proportion to the power and volume of his production; that the necessary consequence of competition is the accumulation of capital in a few hands and hence the restoration of monopoly in a more terrible form; and that, finally, the distinction between capitalist and landlord, between agricultural worker and industrial worker, disappears and the whole of society must split into the two classes of *property owners* and propertyless *workers.*

Political economy proceeds from the fact of private property. It does not explain it. It grasps the *material* process of private property, the process through which it actually passes, in general and abstract formulae which it then takes as *laws*. It does not *comprehend* these laws – i.e., it does not show how they arise from the nature of private property. Political economy fails to explain the reason for the division between labour and capital. For example, when it defines the relation of wages to profit, it takes the interests of the capitalists as the basis of its analysis – i.e., it assumes what it is supposed to explain. Similarly, competition is frequently brought into the argument and explained in terms of external circumstances. Political economy teaches us nothing about the extent to which these external and apparently accidental circumstances are only the expression of a necessary development. We have seen how exchange itself appears to political economy as an accidental fact. The only wheels which political economy sets in motion are *greed*, and the *war of the avaricious – Competition.*

Precisely because political economy fails to grasp the interconnections within the movement, it was possible to oppose, for example, the doctrine of competition to the doctrine of monopoly, the doctrine of craft freedom to the doctrine of the guild, and the doctrine of the division of landed property to the doctrine of the great estate; for competition, craft freedom, and division of landed property were developed and conceived only as accidental, deliberate, violent consequences of monopoly, of the guilds, and of feudal property, and not as their necessary, inevitable, and natural consequences.

We now have to grasp the essential connection between private property, greed, the separation of labour, capital and landed property, exchange and competition, value and the devaluation of man, monopoly, and competition, etc. – the connection between this entire system of estrangement and the *money* system.

We must avoid repeating the mistake of the political economist, who bases his explanations on some imaginary primordial condition. Such a primordial condition explains nothing. It simply pushes the question into the grey and nebulous distance. It assumes as facts and events what it is supposed to deduce – namely, the necessary relationships between two things, between, for example, the division of labour and exchange. Similarly, theology explains the origin of evil by the fall of Man – i.e., it assumes as a fact in the form of history what it should explain.

We shall start out from an *actual* economic fact.

The worker becomes poorer the more wealth he produces, the more his production increases in power and extent. The worker becomes an ever cheaper commodity the more commodities he produces. The *devaluation* of the human world grows in direct proportion to the *increase in value* of the world of things. Labour not only produces commodities; it also produces itself and the workers as a *commodity* and it does so in the same proportion in which it produces commodities in general.

This fact simply means that the object that labour produces, its product, stands opposed to it as *something alien*, as a power independent of the producer. The product of labour is labour embodied and made material in an object, it is the *objectification* of labour. The realization of labour is its objectification. In the sphere of political economy, this realization of labour appears as a *loss of reality* for the worker,[1] objectification as loss of and bondage to the object, and appropriation as estrangement, as *alienation*.[2]

So much does the realization of labour appear as loss of reality that the worker loses his reality to the point of dying of starvation. So much does objectification appear as loss of the object that the worker is robbed of the objects he needs most not only for life but also for work. Work itself becomes an object which he can only obtain through an enormous effort and with spasmodic interruptions. So much does the appropriation of the object appear as estrangement that the more objects the worker produces the fewer can he possess and the more he falls under the domination of his product, of capital.

All these consequences are contained in this characteristic, that the worker is related to the *product of labour* as to an *alien* object. For it is clear that, according to this premise, the more the worker exerts himself in his work, the more powerful the alien, objective world becomes which he brings into being over against himself, the poorer he and his inner world become, and the less they belong to him. It is the same in religion. The more man puts into God, the less he retains within himself. The worker places his life in the object; but now it no longer belongs to him, but to the object. The greater his activity, therefore, the fewer objects the worker possesses. What the product of his labour is, he is not. Therefore, the greater this product, the less is he himself. The externalisation of the worker in his product means not only that his labour becomes an object, an *external* existence, but that it exists *outside him*, independently of him and alien to him, and begins to confront him as an autonomous power; that the life which he has bestowed on the object confronts him as hostile and alien.

Let us now take a closer look at objectification, at the production of the worker, and the estrangement, the loss of the object, of his product, that this entails.

The workers can create nothing without nature, without the sensuous external world. It is the material in which his labour realizes itself, in which it is active and from which, and by means of which, it produces.

But just as nature provides labour with the means of life, in the sense of labour cannot live without objects on which to exercise itself, so also it provides the means of life in the narrower sense, namely the means of physical subsistence of the worker.

The more the worker appropriates the external world, sensuous nature, through his labour, the more he deprives himself of the means of life in two respects: firstly, the sensuous external world becomes less and less an object belonging to his labour, a means of life of his labour; and, secondly, it becomes less and less a means of life in the immediate sense, a means for the physical subsistence of the worker.

In these two respects, then, the worker becomes a slave of his object; firstly, in that he receives an object of labour, i.e., he receives work, and, secondly, in that he receives means of subsistence. Firstly, then, so that he can exist as a worker, and secondly as a physical subject. The culmination of this slavery is that it is only as a worker that he can maintain himself as a physical subject and only as a physical subject that he is a worker.

(The estrangement of the worker in his object is expressed according to the laws of political economy in the following way:

1. the more the worker produces, the less he has to consume;
2. the more value he creates, the more worthless he becomes;
3. the more his product is shaped, the more misshapen the worker;
4. the more civilized his object, the more barbarous the worker;
5. the more powerful the work, the more powerless the worker;
6. the more intelligent the work, the duller the worker and the more he becomes a slave of nature.)

Political economy conceals the estrangement in the nature of labour by ignoring the direct relationship between the worker (labour) *and production.* It is true that labour produces marvels for the rich, but it produces privation for the worker. It produces palaces, but hovels for the worker. It produces beauty, but deformity for the worker. It replaces labour by machines, but it casts some of the workers back into barbarous forms of labour and turns others into machines. It produces intelligence, but it produces idiocy and cretinism for the worker.

The direct relationship of labour to its products is the relationship of the worker to the objects of his production. The relationship of the rich man to the objects of production and to production itself is only a *consequence* of this first relationship, and confirms it. Later, we shall consider this second aspect. Therefore, when we ask what is the essential relationship of labour, we are asking about the relationship of the worker to production.

Up to now, we have considered the estrangement, the alienation of the worker, only from one aspect – i.e., *the worker's relationship to the products of his labour*. But estrangement manifests itself not only in the result, but also in the *act of production*, within the *activity of production* itself. How could the product of the worker's activity confront him as something alien if it were not for the fact that in the act of production he was estranging himself from himself? After all, the product is simply the resumé of the activity, of the production. So if the product of labour is alienation, production itself must be active alienation, the alienation of activity, the activity of alienation. The estrangement of the object of labour merely summarizes the estrangement, the alienation in the activity of labour itself.

What constitutes the alienation of labour?

Firstly, the fact that labour is *external* to the worker – i.e., does not belong to his essential being; that he, therefore, does not confirm himself in his work, but denies himself, feels miserable and not happy, does not develop free mental and physical energy, but mortifies his flesh and ruins his mind. Hence, the worker feels himself only when he is not working; when he is working, he does not feel himself. He is at home when he is not working, and not at home when he is working. His labour is, therefore, not voluntary but forced, it is *forced labour*. It is, therefore, not the satisfaction of a need but a mere *means* to satisfy needs outside itself. Its alien character is clearly demonstrated by the fact that as soon as no physical or other compulsion exists, it is shunned like the plague. External labour, labour in which man alienates himself, is a labour of self-sacrifice, of mortification. Finally, the external character of labour for the worker is demonstrated by the fact that it belongs not to him but to another, and that in it he belongs not to himself but to another. Just as in religion the spontaneous activity of the human imagination, the human brain, and the human heart, detaches itself from the individual and reappears as the alien activity of a god or of a devil, so the activity of the worker is not his own spontaneous activity. It belongs to another, it is a loss of his self.

The result is that man (the worker) feels that he is acting freely only in his animal functions – eating, drinking, and procreating, or at most in his dwelling and adornment – while in his human functions, he is nothing more than animal.

It is true that eating, drinking, and procreating, etc., are also genuine human functions. However, when abstracted from other aspects of human activity, and turned into final and exclusive ends, they are animal.

We have considered the act of estrangement of practical human activity, of labour, from two aspects:

(1) the relationship of the worker to the product of labour as an alien object that has power over him. The relationship is, at the same time, the relationship to the sensuous external world, to natural objects, as an alien world confronting him, in hostile opposition.

(2) The relationship of labour to the *act of production* within labour. This relationship is the relationship of the worker to his own activity as something which is alien and does not belong to him, activity as passivity, power as impotence, procreation as emasculation, the worker's own physical and mental energy, his personal life – for what is life but activity? – as an activity directed against himself, which is independent of him and does not belong to him. Self-estrangement, as compared with the estrangement of the object mentioned above.

We now have to derive a third feature of estranged labour from the two we have already examined.

Man is a species-being,[3] not only because he practically and theoretically makes the species – both his own and those of other things – his object, but also – and this is simply another way of saying the same thing – because he looks upon himself as the present, living species, because he looks upon himself as a universal and therefore free being.

Species-life, both for man and for animals, consists physically in the fact that man, like animals, lives from inorganic nature; and because man is more universal than animals, so too is the area of inorganic nature from which he lives more universal. Just as plants, animals, stones, air, light, etc., theoretically form a part of human consciousness, partly as objects of science and partly as objects of art – his spiritual inorganic nature, his spiritual means of life, which he must first prepare before he can enjoy and digest them – so, too, in practice they form a part of human life and human activity. In a physical sense, man lives only from these natural products, whether in the form of nourishment, heating, clothing, shelter, etc. The universality of man manifests itself in practice in that universality which makes the whole of nature his inorganic body, (1) as a direct means of life and (2) as the matter, the object, and the tool of his life activity. Nature is man's inorganic body – that is to say, nature insofar as it is not the human body. Man lives from nature – i.e., nature is his body – and he must maintain a continuing dialogue with it is he is not to die. To say that man's physical and mental life is linked to nature simply means that nature is linked to itself, for man is a part of nature.

Estranged labour not only (1) estranges nature from man and (2) estranges man from himself, from his own function, from his vital activity; because of this, it also estranges man from his species. It turns his species-life into a means for his individual life. Firstly, it estranges species-life and individual life, and, secondly, it turns the latter, in its abstract form, into the purpose of the former, also in its abstract and estranged form.

For in the first place labour, life activity, productive life itself, appears to man only as a means for the satisfaction of a need, the need to preserve physical existence. But productive life is species-life. It is life-producing life. The whole character of a species, its species-character, resides in the nature of its life activity, and free conscious activity constitutes the species-character of man. Life appears only as a means of life.

The animal is immediately one with its life activity. It is not distinct from that activity; it is that activity. Man makes his life activity itself an object of his will and consciousness. He has conscious life activity. It is not a determination with which he directly merges. Conscious life activity directly distinguishes man from animal life activity. Only because of that is he a species-being. Or, rather, he is a conscious being – i.e., his own life is an object for him, only because he is a species-being. Only because of that is his activity free activity. Estranged labour reverses the relationship so that man, just because he is a conscious being, makes his life activity, his *essential being*, a mere means for his *existence*.

The practical creation of an *objective world*, the fashioning of inorganic nature, is proof that man is a conscious species-being – i.e., a being which treats the species as its own essential being or itself as a species-being. It is true that animals also produce. They build nests and dwellings, like the bee, the beaver, the ant, etc. But they produce only their own immediate needs or those of their young; they produce only when immediate physical need compels them to do so, while man produces even when he is free from physical need and truly produces only in freedom from such need; they produce only themselves, while man reproduces the whole of nature; their products belong immediately to their physical bodies, while man freely confronts his own product. Animals produce only according to the standards and needs of the species to which they belong, while man is capable of producing according to the standards of every species and of applying to each object its inherent standard; hence, man also produces in accordance with the laws of beauty.

It is, therefore, in his fashioning of the objective that man really proves himself to be a species-being. Such production is his active species-life. Through it, nature appears as *his* work and his reality. The object of labour is, therefore, the objectification of the species-life of man: for man produces himself not only intellectually, in his consciousness, but actively and actually, and he can therefore contemplate himself in a world he himself has created. In tearing away the object of his production from man, estranged labour therefore tears away from him his species-life, his true species-objectivity, and transforms his advantage over animals into the disadvantage that his inorganic body, nature, is taken from him.

In the same way as estranged labour reduces spontaneous and free activity to a means, it makes man's species-life a means of his physical existence.

Consciousness, which man has from his species, is transformed through estrangement so that species-life becomes a means for him.

(3) Estranged labour, therefore, turns man's species-being – both nature and his intellectual species-power – into a being alien to him and a means of his individual existence. It estranges man from his own body, from nature as it exists outside him, from his spiritual essence, his human existence.

(4) An immediate consequence of man's estrangement from the product of his labour, his life activity, his species-being, is the estrangement of man from man. When man confronts himself, he also confronts other men. What is true of man's relationship to his labour, to the product of his labour, and to himself, is also true of his relationship to other men, and to the labour and the object of the labour of other men.

In general, the proposition that man is estranged from his species-being means that each man is estranged from the others and that all are estranged from man's essence.

Man's estrangement, like all relationships of man to himself, is realized and expressed only in man's relationship to other men.

In the relationship of estranged labour, each man therefore regards the other in accordance with the standard and the situation in which he as a worker finds himself.

We started out from an economic fact, the estrangement of the worker and of his production. We gave this fact conceptual form: estranged, alienated labour. We have analyzed this concept, and in so doing merely analyzed an economic fact.

Let us now go on to see how the concept of estranged, alienated labour must express and present itself in reality.

If the product of labour is alien to me, and confronts me as an alien power, to whom does it then belong?

To a being *other* than me.

Who is this being?

The gods? It is true that in early times most production – e.g., temple building, etc., in Egypt, India, and Mexico – was in the service of the gods, just as the product belonged to the gods. But the gods alone were never the masters of labour. The same is true of nature. And what a paradox it would be if the more man subjugates nature through his labour and the more divine miracles are made superfluous by the miracles of industry, the more he is forced to forgo the joy or production and the enjoyment of the product out of deference to these powers.

The alien being to whom labour and the product of labour belong, in whose service labour is performed, and for whose enjoyment the product of labour is created, can be none other than man himself.

If the product of labour does not belong to the worker, and if it confronts him as an alien power, this is only possible because it belongs to a man other than the worker. If his activity is a torment for him, it must provide pleasure and enjoyment for someone else. Not the gods, not nature, but only man himself can be this alien power over men.

Consider the above proposition that the relationship of man to himself becomes objective and real for him only through his relationship to other men. If, therefore, he regards the product of his labour, his objectified labour, as an alien, hostile, and powerful object which is independent of him, then his relationship to that object is such that another man – alien, hostile, powerful, and independent of him – is its master. If he relates to his own activity as unfree activity, then he relates to it as activity in the service, under the rule, coercion, and yoke of another man.

Every self-estrangement of man from himself and nature is manifested in the relationship he sets up between other men and himself and nature. Thus, religious self-estrangement is necessarily manifested in the relationship between layman and priest, or, since we are dealing here with the spiritual world, between layman and mediator, etc. In the practical, real world, self-estrangement can manifest itself only in the practical, real relationship to other men. The medium through which estrangement progresses is itself a practical one. So through estranged labour man not only produces his relationship to the object and to the act of production as to alien and hostile powers; he also produces the relationship in which other men stand to his production and product, and the relationship in which he stands to these other men. Just as he creates his own production as a loss of reality, a punishment, and his own product as a loss, a product which does not belong to him, so he creates the domination of the non-producer over production and its product. Just as he estranges from himself his own activity, so he confers upon the stranger and activity which does not belong to him.

Up to now, we have considered the relationship only from the side of the worker. Later on, we shall consider it from the side of the non-worker.

Thus, through estranged, alienated labour, the worker creates the relationship of another man, who is alien to labour and stands outside it, to that labour. The relation of the worker to labour creates the relation of the capitalist – or whatever other word one chooses for the master of labour – to that labour. Private property is therefore the product, result, and necessary consequence of alienated labour, of the external relation of the worker to nature and to himself.

Private property thus derives from an analysis of the concept of alienated labour – i.e., alienated man, estranged labour, estranged life, estranged man.

It is true that we took the concept of alienated labour (alienated life) from political economy as a result of the movement of private property. But it is clear from an analysis of this concept that, although private property appears as the basis and cause of alienated labour, it is in fact its consequence, just as the gods were originally not the cause but the effect of the confusion in men's minds. Later, however, this relationship becomes reciprocal.

It is only when the development of private property reaches its ultimate point of culmination that this, its secret, re-emerges; namely, that is

(a) the product of alienated labour, and
(b) the means through which labour is alienated, the realization of this alienation.

This development throws light upon a number of hitherto unresolved controversies.

(1) Political economy starts out from labour as the real soul of production and yet gives nothing to labour and everything to private property. Proudhon has dealt with this contradiction by deciding for labour and against private property.[4] But we have seen that this apparent contradiction is the contradiction of *estranged labour* with itself and that political economy has merely formulated laws of estranged labour.

It, therefore, follows for us that wages and private property are identical: for there the product, the object of labour, pays for the labour itself, wages are only a necessary consequence of the estrangement of labour; similarly, where wages are concerned, labour appears not as an end in itself but as the servant of wages. We intend to deal with this point in more detail later on: for the present we shall merely draw a few conclusions.[5]

An enforced rise in wages (disregarding all other difficulties, including the fact that such an anomalous situation could only be prolonged by force) would therefore be nothing more than better pay for slaves and would not mean an increase in human significance or dignity for either the worker or the labour.

Even the equality of wages, which Proudhon demands, would merely transform the relation of the present-day worker to his work into the relation of all men to work. Society would then be conceived as an abstract capitalist.

Wages are an immediate consequence of estranged labour, and estranged labour is the immediate cause of private property. If the one falls, then the other must fall too.

(2) It further follows from the relation of estranged labour to private property that the emancipation of society from private property, etc., from servitude, is expressed in the *political* form of the *emancipation of the workers.* This is not because it is only a question of *their* emancipation, but because in their emancipation is contained universal human emancipation. The reason for this universality is that the whole of human servitude is involved in the relation of the worker to production, and all relations of servitude are nothing but modifications and consequences of this relation.

Just as we have arrived at the concept of *private property* through an analysis of the concept of *estranged, alienated labour*, so with the help of these two factors it is possible to evolve all economic categories, and in each of these categories – e.g., trade, competition, capital, money – we shall identify only a particular and developed expression of these basic constituents.

But, before we go on to consider this configuration, let us try to solve two further problems.

(1) We have to determine the general nature of private property, as it has arisen out of estranged labour, in its relation to truly human and social property.

(2) We have taken the *estrangement of labour*, its *alienation*, as a fact and we have analyzed that fact. How, we now ask, does *man* come to *alienate* his labour, to estrange it? How is this estrangement founded in the nature of human development? We have already gone a long way towards solving this problem by *transforming* the question of the *origin of private property* into the question of the relationship of alienated labour to the course of human development. For, in speaking of private property, one imagines that one is dealing with something external to man. In speaking of labour, one is dealing immediately with man himself. This new way of formulating the problem already contains its solution.

ad (1): *The general nature of private property and its relationship to truly human property.*

Alienated labour has resolved itself for us into two component parts, which mutually condition one another, or which are merely different expressions of one and the same relationship. Appropriation appears as *estrangement*, as *alienation*; and *alienation* appears as *appropriation*, estrangement as true *admission to citizenship*.[6]

We have considered the one aspect – alienated labour in relation to the worker himself – i.e., *the relation of alienated labour to itself.* And as product, as necessary consequence of this relationship, we have found the property relation of the non-worker to the worker and to labour. Private property as the material, summarized expression of alienated labour embraces both relations – the relation of the worker to labour and to the product of his labour and the non-workers, and the relation of the non-worker to the worker and to the product of his labour.

We have already seen that, in relation to the worker who appropriates nature through his labour, appropriation appears as estrangement, self-activity as activity for another and of another, vitality as a sacrifice of life, production of an object as loss of that object to an alien power, to an *alien* man. Let us now consider the relation between this man, who is *alien* to labour and to the worker, and the worker, labour, and the object of labour.

The first thing to point out is that everything which appears for the worker as an activity of alienation, of estrangement, appears for the non-worker as a situation of alienation, of estrangement.

Secondly, the real, practical attitude of the worker in production and to the product (as a state of mind) appears for the non-worker who confronts him as a theoretical attitude.

Thirdly, the non-worker does everything against the worker which the worker does against himself, but he does not do against himself what he does against the worker.

Let us take a closer look at these three relationships.

From the "Preface" to *A Contribution to the Critique of Political Economy*

In the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely relations of production appropriate to a given stage in the development of their material forces of production. The totality of these relations of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The mode of production of material life conditions the general process of social, political and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. At a certain stage of development, the material productive forces of society come into conflict with the existing relations of production or – this merely expresses the same thing in legal terms – with the property relations within the framework of which they have operated hitherto. From forms of development of the productive forces these relations turn into their fetters. Then begins an era of social revolution. The changes in the economic foundation lead sooner or later to the transformation of the whole immense superstructure.

In studying such transformations it is always necessary to distinguish between the material transformation of the economic conditions of production, which can be determined with the precision of natural science, and the legal, political, religious, artistic or philosophic – in short, ideological forms in which men become conscious of this conflict and fight it out. Just as one does not judge an individual by what he thinks about himself, so one cannot judge such a period of transformation by its consciousness, but, on the contrary, this consciousness must be explained from the contradictions of material life, from the conflict existing between the social forces of production and the relations of production. No social order is ever destroyed before all the productive forces for which it is sufficient have been developed, and new superior relations of production never replace older ones before the material conditions for their existence have matured within the framework of the old society.

Mankind thus inevitably sets itself only such tasks as it is able to solve, since closer examination will always show that the problem itself arises only when the material conditions for its solution are already present or at least in the course of formation. In broad outline, the Asiatic, ancient, feudal and modern bourgeois modes of production may be designated as epochs marking progress in the economic development of society. The bourgeois mode of production is the last antagonistic form of the social process of production – antagonistic not in the sense of individual antagonism but of an antagonism that emanates from the individuals' social conditions of existence – but the productive forces developing within bourgeois society create also the material conditions for a solution of this antagonism. The prehistory of human society accordingly closes with this social formation.

The Method of Political Economy

When examining a given country from the standpoint of political economy, we begin with its population, the division of the population into classes, town and country, the sea, the different branches of production, export and import, annual production and consumption, prices, etc.

It would seem to be the proper thing to start with the real and concrete elements, with the actual preconditions, *e.g.*, to start in the sphere of economy with population, which forms the basis and the subject of the whole social process of production. Closer consideration shows, however, that this is wrong. Population is an abstraction if, for instance, one disregards the classes of which it is composed. These classes in turn remain empty terms if one does not know the factors on which they depend, *e.g.*, wage-labour, capital, and so on. These presuppose exchange, division of labour, prices, etc. For example, capital is nothing without wage-labour, without value, money, price, etc. If one were to take population as the point of departure, it would be a very vague notion of a complex whole and through closer definition one would arrive analytically at increasingly simple concepts; from imaginary concrete terms one would move to more and more tenuous abstractions until one reached the most simple definitions. From there it would be necessary to make the journey again in the opposite direction until one arrived once more at the concept of population, which is this time not a vague notion of a whole, but a totality comprising many determinations and relations. The first course is the historical one taken by political economy at its inception. The seventeenth-century economists, for example, always took as their starting point the living organism, the population, the nation, the State, several States, etc., but analysis led them always in the end to the discovery of a few decisive abstract, general relations, such as division of labour, money, and value. When these separate factors were more or less clearly deduced and established, economic systems were evolved which from simple concepts, such as labour, division of labour, demand, exchange-value, advanced to categories like State, international exchange and world market. The latter is obviously the correct scientific method. The concrete concept is concrete because it is a synthesis of many definitions, thus representing the unity of diverse aspects. It appears therefore in reasoning as a summing-up, a result, and not as the starting point, although it is the real point of origin, and thus also the point of origin of perception and imagination. The first procedure attenuates meaningful images to abstract definitions, the second leads from abstract definitions by way of reasoning to the reproduction of the concrete situation. Hegel accordingly conceived the illusory idea that the real world is the result of thinking which causes its own synthesis, its own deepening and its own movement; whereas the method of advancing from the abstract to the concrete is simply the way in which thinking assimilates the concrete and reproduces it as a concrete mental category. This is, however, by no means the process of evolution of the concrete world itself. For example, the simplest economic category, *e.g.*, exchange-value, presupposes population, a population moreover which produces under definite conditions, as well as a distinct kind of family, or community, or State, etc. Exchange-value cannot exist except as an abstract, *unilateral* relation of an already existing concrete organic whole. But exchange-value as a category leads an antediluvian existence. Thus to consciousness-and this comprises philosophical consciousness – which regards the comprehending mind as the real man, and hence the comprehended world as such as the only real world; to consciousness, therefore, the evolution of categories appears as the actual process of production – which unfortunately is given an impulse from outside – whose result is the world; and this (which is however again a tautological expression) is true in so far as the concrete totality regarded as a conceptual totality, as a mental fact, is indeed a product of thinking, of comprehension; but it is by no means a product of the idea which evolves spontaneously and whose thinking proceeds outside and above perception and imagination, but is the result of the assimilation and transformation of perceptions and images into concepts. The totality as a conceptual entity seen by the intellect is a product of the thinking intellect which assimilates the world in the only way open to it, a way which differs from the artistic, religious and practically intelligent assimilation of this world. The concrete subject remains outside the intellect and independent of it – that is so long as the intellect adopts a purely speculative, purely theoretical attitude. The subject, society, must always be envisaged therefore as the pre-condition of comprehension even when the theoretical method is employed.

But have not these simple categories also an independent historical or natural existence preceding that of the more concrete ones? This depends. Hegel, for example, correctly takes ownership, the simplest legal relation of the subject, as the point of departure of the philosophy of law. No ownership exists, however, before the family or the relations of master and servant are evolved, and these are much more concrete relations. It would, on the other hand, be correct to say that families and entire tribes exist which have as yet only *possessions* and not *property*. The simpler category appears thus as a relation of simple family or tribal communities to property. In societies which have reached a higher stage the category appears as a comparatively simple relation existing in a more advanced community. The concrete substratum underlying the relation of ownership is however always presupposed. One can conceive an individual savage who has possessions; possession in this case, however, is not a legal relation. It is incorrect that in the course of historical development possession gave rise to the family. On the contrary, possession always presupposes this "more concrete legal category." One may, nevertheless, conclude that the simple categories represent relations or conditions which may reflect the immature concrete situation without as yet positing the more complex relation or condition which is conceptually expressed in the more concrete category; on the other hand, the same category may be retained as a subordinate relation in more developed concrete circumstances. Money may exist and has existed in historical time before capital, banks, wage-labour, etc. came into being. In this respect it can be said, therefore, that the simpler category expresses relations predominating in an immature entity or subordinate relations in a more advanced entity; relations which already existed historically before the entity had developed the aspects expressed in a more concrete category. The procedure of abstract reasoning which advances from the simplest to more complex concepts to that extent conforms to actual historical development.

It is true, on the other hand, that there are certain highly developed, but nevertheless historically immature, social formations which employ some of the most advanced economic forms, *e.g.*, cooperation, developed division of labour, etc., without having developed any money at all, for instance Peru. In Slavonic communities too, money – and its pre-condition, exchange – is of little or no importance within the individual community, but is used on the borders, where commerce with other communities takes place; and it is altogether wrong to assume that exchange within the community is an original constituent element. On the contrary, in the beginning exchange tends to arise in the intercourse of different communities with one another, rather than among members of the same community. Moreover, although money begins to play a considerable role very early and in diverse ways, it is known to have been a dominant factor in antiquity only among nations developed in a particular direction, i.e., merchant nations. Even among the Greeks and Romans, the most advanced nations of antiquity, money reaches its full development, which is presupposed in modern bourgeois society, only in the period of their disintegration. Thus the full potential of this quite simple category does not emerge historically in the most advanced phases of society, and it certainly does not penetrate into all economic relations. For example, taxes in kind and deliveries in kind remained the basis of the Roman empire even at the height of its development; indeed a completely evolved monetary system existed in Rome only in the army, and it never permeated the whole complex of labour. Although the simpler category, therefore, may have existed historically before the more concrete category, its complete intensive and extensive development can nevertheless occur in a complex social formation, whereas the more concrete category may have been fully evolved in a more primitive social formation.

Labour seems to be a very simple category. The notion of labour in this universal form, as labour in general, is also extremely old. Nevertheless "labour" in this simplicity is economically considered just as modern a category as the relations which give rise to this simple abstraction. The Monetary System, for example, still regards wealth quite objectively as a thing existing independently in the shape of money. Compared with this standpoint, it was a substantial advance when the Manufacturing or Mercantile System transferred the source of wealth from the object to the subjective activity – mercantile or industrial labour – but it still considered that only this circumscribed activity itself produced money. In contrast to this system, the Physiocrats assume that a specific form of labour – agriculture – creates wealth, and they see the object no longer in the guise of money, but as a product in general, as the universal result of labour. In accordance with the still circumscribed activity, the product remains a naturally developed product, an agricultural product, a product of the land *par excellence.*

It was an immense advance when Adam Smith rejected all restrictions with regard to the activity that produces wealth – for him it was labour as such, neither manufacturing, nor commercial, nor agricultural labour, but all types of labour. The abstract universality which creates wealth implies also the universality of the objects defined as wealth: they are products as such, or once more labour as such, but in this case past, materialised labour. How difficult and immense a transition this was is demonstrated by the fact that Adam Smith himself occasionally relapses once more into the Physiocratic system. It might seem that in this way merely an abstract expression was found for the simplest and most ancient relation in which human beings act as producers – irrespective of the type of society they live in. This is true in one respect, but not in another.

The fact that the specific kind of labour is irrelevant presupposes a highly developed complex of actually existing kinds of labour, none of which is any more the all-important one. The most general abstractions arise on the whole only when concrete development is most profuse, so that a specific quality is seen to be common to many phenomena, or common to all. Then it is no longer perceived solely in a particular form. This abstraction of labour is, on the other hand, by no means simply the conceptual resultant of a variety of concrete types of labour. The fact that the particular kind of labour employed is immaterial is appropriate to a form of society in which individuals easily pass from one type of labour to another, the particular type of labour being accidental to them and therefore irrelevant. Labour, not only as a category but in reality, has become a means to create wealth in general, and has ceased to be tied as an attribute to a particular individual. This state of affairs is most pronounced in the United States, the most modern form of bourgeois society. The abstract category "labour," "labour as such," labour *sans phrase*, the point of departure of modern economics, thus becomes a practical fact only there. The simplest abstraction, which plays a decisive role in modem political economy, an abstraction which expresses an ancient relation existing in all social formations, nevertheless appears to be actually true in this abstract form only as a category of the most modern society. It might be said that phenomena which are historical products in the United States – *e.g.*, the irrelevance of the particular type of labour – appear to be among the Russians, for instance, naturally developed predispositions. But in the first place, there is an enormous difference between barbarians having a predisposition which makes it possible to employ them in various tasks, and civilised people who apply themselves to various tasks. As regards the Russians, moreover, their indifference to the particular kind of labour performed is in practice matched by their traditional habit of clinging fast to a very definite kind of labour from which they are extricated only by external influences.

The example of labour strikingly demonstrates how even the most abstract categories, despite their validity in all epochs – precisely because they are abstractions – are equally a product of historical conditions even in the specific form of abstractions, and they retain their full validity only for and within the framework of these conditions.

Bourgeois society is the most advanced and complex historical organisation of production. The categories which express its relations, and an understanding of its structure, therefore, provide an insight into the structure and the relations of production of all formerly existing social formations the ruins and component elements of which were used in the creation of bourgeois society. Some of these unassimilated remains are still carried on within bourgeois society, others, however, which previously existed only in rudimentary form, have been further developed and have attained their full significance, etc. The anatomy of man is a key to the anatomy of the ape. On the other hand, rudiments of more advanced forms in the lower species of animals can only be understood when the more advanced forms are already known. Bourgeois economy thus provides a key to the economy of antiquity, etc. But it is quite impossible (to gain this insight) in the manner of those economists who obliterate all historical differences and who see in all social phenomena only bourgeois phenomena. If one knows rent, it is possible to understand tribute, tithe, etc., but they do not have to be treated as identical.

Since bourgeois society is, moreover, only a contradictory form of development, it contains relations of earlier societies often merely in very stunted form or even in the form of travesties, *e.g.*, communal ownership. Thus, although it is true that the categories of bourgeois economy are valid for all other social formations, this has to be taken *cum grano salis*, for they may contain them in an advanced, stunted, caricatured, etc., form that is always with substantial differences. What is called historical evolution depends in general on the fact that the latest form regards earlier ones as stages in the development of itself and conceives them always in a one-sided manner, since only rarely and under quite special conditions is a society able to adopt a critical attitude towards itself; in this context we are not of course discussing historical periods which themselves believe that they are periods of decline. The Christian religion was able to contribute to an objective understanding of earlier mythologies only when its self-criticism was to a certain extent prepared, as it were potentially. Similarly, only when the self-criticism of bourgeois society had begun, was bourgeois political economy able to understand the feudal, ancient and oriental economies. In so far as bourgeois political economy did not simply identify itself with the past in a mythological manner, its criticism of earlier economies-especially of the feudal system against which it still had to wage a direct struggle-resembled the criticism that Christianity directed against heathenism, or which Protestantism directed against Catholicism.

Just as in general when examining any historical or social science, so also in the case of the development of economic categories is it always necessary to remember that the subject, in this context contemporary bourgeois society, is presupposed both in reality and in the mind, and that therefore categories express forms of existence and conditions of existence – and sometimes merely separate aspects – of this particular society, the subject; thus the category, *even from the scientific standpoint*, by no means begins at the moment when it is discussed *as such*. This has to be remembered because it provides important criteria for the arrangement of the material. For example, nothing seems more natural than to begin with rent, i.e., with landed property, since it is associated with the earth, the source of all production and all life, and with agriculture, the first form of production in all societies that have attained a measure of stability. But nothing would be more erroneous. There is in every social formation a particular branch of production which determines the position and importance of all the others, and the relations obtaining in this branch accordingly determine the relations of all other branches as well. It is as though light of a particular hue were cast upon everything, tingeing all other colours and modifying their specific features; or as if a special ether determined the specific gravity of everything found in it. Let us take as an example pastoral tribes. (Tribes living exclusively on hunting or fishing are beyond the boundary line from which real development begins.) A certain type of agricultural activity occurs among them and this determines land ownership. It is communal ownership and retains this form in a larger or smaller measure, according to the degree to which these people maintain their traditions, *e.g.*, communal ownership among the Slavs. Among settled agricultural people-settled already to a large extent-where agriculture predominates as in the societies of antiquity and the feudal period, even manufacture, its structure and the forms of property corresponding thereto, have, in some measure, specifically agrarian features. Manufacture is either completely dependent on agriculture, as in the earlier

Roman period, or as in the Middle Ages, it copies in the town and in its conditions the organisation of the countryside. In the Middle Ages even capital – unless it was solely money capital – consisted of the traditional tools, etc., and retained a specifically agrarian character. The reverse takes place in bourgeois society. Agriculture to an increasing extent becomes just a branch of industry and is completely dominated by capital. The same applies to rent. In all forms in which landed property is the decisive factor, natural relations still predominate; in the forms in which the decisive factor is capital, social, historically evolved elements predominate. Rent cannot be understood without capital, but capital can be understood without rent. Capital is the economic power that dominates everything in bourgeois society. It must form both the point of departure and the conclusion and it has to be expounded before landed property. After analysing capital and landed property separately, their interconnection must be examined.

It would be inexpedient and wrong therefore to present the economic categories successively in the order in which they have played the dominant role in history. On the contrary, their order of succession is determined by their mutual relation in modern bourgeois society and this is quite the reverse of what appears to be natural to them or in accordance with the sequence of historical development The point at issue is not the role that various economic relations have played in the succession of various social formations appearing in the course of history; even less is it their sequence "as concepts" (*Proudhon*) (a nebulous notion of the historical process), but their position within modern bourgeois society.

It is precisely the predominance of agricultural peoples in the ancient world which caused the merchant nations – Phoenicians, Carthaginians – to develop in such purity (abstract precision). For capital in the shape of merchant or money capital appears in that abstract form where capital has not yet become the dominant factor in society. Lombards and Jews occupied the same position with regard to mediaeval agrarian societies.

Another example of the various roles which the same categories have played at different stages of society are joint-stock companies, one of the most recent features of bourgeois society; but they arise also in its early period in the form of large privileged commercial companies with rights of monopoly.

The concept of national wealth finds its way into the works of the economists of the seventeenth century as the notion that wealth is created for the State, whose power, on the other hand, is proportional to this wealth – a notion which to some extent still survives even among eighteenth-century economists. This is still an unintentionally hypocritical manner in which wealth and the production of wealth are proclaimed to be the goal of the modern State, which is regarded merely as a means for producing wealth.

The disposition of material has evidently to be made in such a way that [section] one comprises general abstract definitions, which therefore appertain in some measure to all social formations, but in the sense set forth earlier. Two, the categories which constitute the internal structure of bourgeois society and on which the principal classes are based. Capital, wage-labour, landed property and their relations to one another. Town and country. The three large social classes; exchange between them. Circulation. The (private) credit system. Three, the State as the epitome of bourgeois society. Analysis of its relations to itself. The "unproductive" classes. Taxes. National debt. Public credit. Population. Colonies. Emigration. Four, international conditions of production. International division of labour. International exchange. Export and import. Rate of exchange. Five, world market and crises.

Notes

1. Marx, still using Hegel's terminology and his approach to the unity of the opposites, counterposes the term "Verwirklichung" (realisation) to "Entwirklichung" (loss of realisation).
2. In this manuscript Marx frequently uses two similar German terms, "Entäusserung" and "Entfremdung," to express the notion of "alienation." In the present edition the former is generally translated as "alienation," the latter as "estrangement," because in the later economic works (*Theories of Surplus-Value*) Marx himself used the word "alienation" as the English equivalent of the term "Entäusserung."
3. The term "species-being" (Gattungswesen) is derived from Ludwig Feuerbach's philosophy where it is applied to man and mankind as a whole.
4. Apparently Marx refers to Proudhon's book *Qu'est-ce que la propriété?*, Paris, 1841.
5. This passage shows that Marx here uses the category of wages in a broad sense, as an expression of antagonistic relations between the classes of capitalists and of wage-workers. Under "the wages" he understands "the wage-labour," the capitalist system as such. This idea was apparently elaborated in detail in that part of the manuscript which is now extant.
6. This apparently refers to the conversion of individuals into members of civil society which is considered as the sphere of property, of material relations that determine all other relations. In this case Marx refers to the material relations of society based on private property and the antagonism of different classes.

SIX

The Limitations of Marginal Utility

Thorstein Veblen

Thorstein Veblen (1857–1929) was born in Wisconsin and received his Ph.D. from Yale. He taught at the University of Chicago and at several other schools. Given his irascibility and his radical criticisms of American society, he was unable to find a permanent position. Veblen did, however, gain considerable renown – he was even offered the presidency of the American Economic Association (which he turned down). Veblen was, as in the essay reprinted here, a persistent critic of neoclassical economics. In the last two decades, interest in Veblen's economics and the work of other "institutionalist" economists has increased significantly.

The limitations of the marginal-utility economics are sharp and characteristic. It is from first to last a doctrine of value, and in point of form and method it is a theory of valuation. The whole system, therefore, lies within the theoretical field of distribution, and it has but a secondary bearing on any other economic phenomena than those of distribution – the term being taken in its accepted sense of pecuniary distribution, or distribution in point of ownership. Now and again an attempt is made to extend the use of the principle of marginal utility beyond this range, so as to apply it to questions of production, but hitherto without sensible effect, and necessarily so. The most ingenious and the most promising of such attempts have been those of Mr. Clark, whose work marks the extreme range of endeavor and the extreme degree of success in so seeking to turn a postulate of distribution to account for a theory of production. But the outcome has been a doctrine of the production of values, and value, in Mr. Clark's as in other utility systems, is a matter of valuation; which throws the whole excursion back into the field of distribution. Similarly, as regards attempts to make use of this principle in an analysis of the phenomena of consumption, the best results arrived

From "The Limitation of Marginal Utility," by Thorstein Veblen, *Journal of Political Economy* vol. 17 (1909): 620–36.

at are some formulation of the pecuniary distribution of consumption goods.

Within this limited range marginal-utility theory is of a wholly statistical character. It offers no theory of a movement of any kind, being occupied with the adjustment of values to a given situation. Of this again, no more convincing illustration need be had than is afforded by the work of Mr. Clark, which is not excelled in point of earnestness, perseverance, or insight. For all their use of the term "dynamic," neither Mr. Clark nor any of his associates in this line of research have yet contributed anything at all appreciable to a theory of genesis, growth, sequence, change, process, or the like, in economic life. They have had something to say as to the bearing which given economic changes, accepted as premises, may have on valuation, and so on distribution; but as to the causes or the unfolding sequence of the phenomena of economic life they had nothing to say hitherto; nor can they, since their theory is not drawn in causal terms but in terms of teleology.

In all this the marginal-utility school is substantially at one with the classical economics of the nineteenth century, the difference between the two being that the former is confined within narrower limits and sticks more consistently to its teleological premises. Both are teleological, and neither can consistently admit arguments from cause to effect in the formulation of their main articles of theory. Neither can deal theoretically with phenomena of change, but at the most only with rational adjustment to change which may be supposed to have supervened.

To the modern scientist the phenomena of growth and change are the most obstrusive and most consequential facts observable in economic life. For an understanding of modern economic life the technological advance of the past two centuries – *e.g.*, the growth of the industrial arts – is of the first importance; but marginal-utility theory does not bear on this matter, nor does this matter bear on marginal-utility theory. As a means of theoretically accounting for this technological movement in the past or in the present, or even as a means of formally, technically stating it as an element in the current economic situation, that doctrine and all its works are altogether idle. The like is true for the sequence of change that is going forward in the pecuniary relations of modern life; the hedonistic postulate and its propositions of differential utility neither have served nor can serve an inquiry into these phenomena of growth, although the whole body of marginal-utility economics lies with the range of these pecuniary phenomena. It has nothing to say to the growth of business usages and expedients or to the concomitant changes in the principles of conduct which govern the pecuniary relations of men, which condition and are conditioned by these altered relations of business life or which bring them to pass.

It is characteristic of the school that whenever an element of the cultural fabric, an institution or any institutional phenomenon, is involved in the facts with which the theory is occupied, such institutional facts are taken for granted, denied, or explained away. If it is a question of price, there is offered an explanation of how exchanges may take place with such effect as to leave money and price out of the account. If it is a question of credit, the effect of credit extension on business traffic is left on one side and there is an explanation of how the borrower and lender coöperate to smooth out their respective income streams of consumable goods or sensations of consumption. The failure of the school in this respect is consistent and comprehensive. And yet these economists are lacking neither in intelligence nor in information. They are, indeed, to be credited, commonly, with a wide range of information and an exact control of materials, as well as with a very alert interest in what is going on; and apart from their theoretical pronouncements the members of the school habitually profess the sanest and most intelligent views of current practical questions, even when these questions touch matters of institutional growth and decay.

The infirmity of this theoretical scheme lies in its postulates, which confine the inquiry to generalisations of the teleological or "deductive" order. These postulates, together with the point of view and logical method that follow from them, the marginal-utility school shares with other economists of the classical line – for this school is but a branch or derivative of the English classical economists of the nineteenth century. The substantial difference between this school and the generality of classical economists lies mainly in the fact that in the marginal-utility economics the common postulates are more consistently adhered to at the same time that they are more neatly defined and their limitations are more adequately realized. Both the classical school in general and its specialized variant, the marginal-utility school, in particular, take as their common point of departure the traditional psychology of the early nineteenth-century hedonists, which is accepted as a matter of course or of common notoriety and is held quite uncritically. The central and well-defined tenet so held is that of the hedonistic calculus. Under the guidance of this tenet and of the other psychological conceptions associated and consonant with it, human conduct is conceived of and interpreted as a rational response to the exigencies of the situation in which mankind is placed; as regards economic conduct it is such a rational and unprejudiced response to the stimulus of anticipated pleasure and pain – being typically and in the main, a response to the promptings of anticipated pleasure, for the hedonists of the nineteenth century and of the marginal-utility school are in the main of an optimistic temper.[1] Mankind is, on the whole and normally, (conceived to be) clearsighted and farsighted in its appreciation of future sensuous gains and losses, although there may be some (inconsiderable) difference between men in this respect. Men's activities differ, therefore, (inconsiderably) in respect of the alertness of the response and the nicety of adjustment of irksome pain-cost to apprehended future sensuous gain; but, in the whole, no other ground or line or guidance of conduct than this rationalistic calculus falls properly within the cognizance of the economic hedonists. Such a theory can take account of conduct only in so far as it is rational conduct, guided by deliberate and exhaustively intelligent choice – wise adaptation to the demands of the main chance.

The external circumstances which condition conduct are variable, of course, and so they will have a varying effect upon conduct; but their variation is, in effect, construed to be of such a character only as to vary the degree of strain to which the human agent is subject by contact with these external circumstances. The cultural elements involved in the theoretical scheme, elements that are of the nature of institutions, human relations governed by use and wont in whatever kind and connection, are not subject to inquiry but are taken for granted as pre-existing in a finished, typical form and as making up a normal and definitive economic situation, under which and in terms of which human intercourse is necessarily carried on. This cultural situation comprises a few large and simple articles of institutional furniture, together with their logical implications or corollaries; but it includes nothing of the consequences or effects caused by these institutional elements. The cultural elements so tacitly postulated as immutable conditions precedent to economic life are ownership and free contract, together with such other features of the scheme of natural rights as are implied in the exercise of these. These cultural products are, for the purpose of the theory, conceived to be given a priori in unmitigated force. They are part of the nature of things; so that there is no need of accounting for them or inquiring into them, as to how they have come to be such as they are, or how and why they have changed and are changing, or what effect all this may have on the relations of men who live by or under this cultural situation.

Evidently the acceptance of these immutable premises, tacitly, because uncritically and as a matter of course, by hedonistic economics gives the science a distinctive character and places it in contrast with other sciences whose premises are of a different order. As has already been indicated, the premises in question, so far as they are peculiar to the hedonistic economics, are (*a*) a certain institutional situation, the substantial feature of which is the natural right of ownership, and (*b*) the hedonistic calculus. The distinctive character given to this system of theory by these postulates and by the point of view resulting from their acceptance may be summed up broadly and concisely in saying that the theory is confined to the ground of sufficient reason instead of proceeding on the ground of efficient cause. The contrary is true of modern science, generally (except mathematics), particularly of such sciences as have to do with the phenomena of life and growth. The difference may seem trivial. It is serious only in its consequences. The two methods of inference – from sufficient reason and from efficient cause – are out of touch with one another and there is no transition from one to the other: no method of converting the procedure or the results of the one into those of the other. The immediate consequence is that the resulting economic theory is of a teleological character – "deductive" or "a priori" as it is often called – instead of being drawn in terms of cause and effect. The relation sought by this theory among the facts with which it is occupied is the control exercised by future (apprehended) events over present conduct. Current phenomena are dealt with as conditioned by their future consequences; and in strict marginal-utility theory they can be dealt with only in respect of their control of the present by consideration of the future. Such a (logical) relation of control or guidance between the future and the present of course involves an exercise of intelligence, a taking thought, and hence an intelligent agent through whose discriminating forethought the apprehended future may affect the current course of events; unless, indeed, one were to admit something in the way of a providential order of nature or some occult line of stress of the nature of sympathetic magic. Barring magical and providential elements, the relation of sufficient reason runs by way of the interested discrimination, the forethought, of an agent who takes thought of the future and guides his present activity by regard for this future. The relation of sufficient reason runs only from the (apprehended) future into the present, and it is solely of an intellectual, subjective, personal, teleological character and force; while the relation of cause and effect runs only in the contrary direction, and it is solely of an objective, impersonal, materialistic character and force. The modern scheme of knowledge, on the whole, rests, for its definitive ground, on the relation of cause and effect; the relation of sufficient reason being admitted only provisionally and as a proximate factor in the analysis, always with the unambiguous reservation that the analysis must ultimately come to rest in terms of cause and effect. The merits of this scientific animus, of course, do not concern the present argument.

Now, it happens that the relation of sufficient reason enters very substantially into human conduct. It is this element of discriminating forethought that distinguishes human conduct from brute behavior. And since the economist's subject of inquiry is this human conduct, that relation necessarily comes in for a large share of his attention in any theoretical formulation of economic phenomena, whether hedonistic or otherwise. But while modern science at large has made the causal relation the sole ultimate ground of theoretical formulation; and while the other sciences that deal with human life admit the relation of sufficient reason as a proximate, supplementary, or intermediate ground, subsidiary, and subservient to the argument from cause to effect; economics has had the misfortune – as seen from the scientific point of view – to let the former supplant the latter. It is, of course, true that human conduct is distinguished from other natural phenomena by the human faculty for taking thought, and any science that has to do with human conduct must face the patent fact that the details of such conduct consequently fall into the teleological form; but it is the peculiarity of the hedonistic economics that by force of its postulates its attention is confined to this teleological bearing of conduct alone. It deals with this conduct only in so far as it may be construed in rationalistic, teleological terms of calculation and choice. But it is at the same time no less true that human conduct, economic or otherwise, is subject to the sequence of cause and effect, by force of such elements as habituation and conventional requirements. But facts of this order, which are to modern science of graver interest than the teleological details of conduct, necessarily fall outside the attention of the hedonistic economist, because they cannot be construed in terms of sufficient reason, such as his postulates demand, or be fitted into a scheme of teleological doctrines.

There is, therefore, no call to impugn these premises of the marginal-utility economics within their field. They commend themselves to all serious and uncritical persons at the first glance. They are principles of action which underlie the current, business-like scheme of economic life, and as such, as practical grounds of conduct, they are not to be called in question without questioning the existing law and order. As a matter of course, men order their lives by these principles and, practically, entertain no question of their stability and finality. That is what is meant by calling them institutions; they are settled habits of thought common to the generality of men. But it would be mere absentmindedness in any student of civilization therefore to admit that these or any other human institutions have this stability which is currently imputed to them or that they are in this way intrinsic to the nature of things. The acceptance by the economists of these or other institutional elements as given and immutable limits their inquiry in a particular and decisive way. It shuts off the inquiry at the point where the modern scientific interest sets in. The institutions in question are no doubt good for their purpose as institutions, but they are not good as premises for a scientific inquiry into the nature, origin, growth, and effects of these institutions and of the mutations which they undergo and which they bring to pass in the community's scheme of life.

To any modern scientist interested in economic phenomena, the chain of cause and effect in which any given phase of human culture is involved, as well as the cumulative changes wrought in the fabric of human conduct itself by the habitual activity of mankind, are matters of more engrossing and more abiding interest than the method of inference by which an individual is presumed invariably to balance pleasure and pain under given conditions that are presumed to be normal and invariable. The former are questions of the life-history of the race or the community, questions of cultural growth and of the fortunes of generations; while the latter is a question of individual casuistry in the face of a given situation that may arise in the course of this cultural growth. The former bear on the continuity and mutations of that scheme of conduct whereby mankind deals with its material means of life; the latter, if it is conceived in hedonistic terms, concerns a disconnected episode in the sensuous experience of an individual member of such a community.

In so far as modern science inquires into the phenomena of life, whether inanimate, brute, or human, it is occupied about questions of genesis and cumulative change, and it converges upon a theoretical formulation in the shape of a life-history drawn in causal terms. In so far as it is a science in the current sense of the term, any science, such as economics, which has to do with human conduct, becomes a genetic inquiry into the human scheme of life; and where, as in economics, the subject of inquiry is the conduct of man in his dealings with the material means of life, the science is necessarily an inquiry into the life-history of material civilization, on a more or less extended or restricted plan. Not that the economist's inquiry isolates material civilization from all other phases and bearings of human culture, and so studies the motions of an abstractly conceived "economic man." On the contrary, no theoretical inquiry into this material civilization that shall be at all adequate to any scientific purpose can be carried out without taking this material civilization in its causal, that is to say, its genetic, relations to other phases and bearings of the cultural complex; without studying it as it is wrought upon by other lines of cultural growth and as working its effects in these other lines. But in so far as the inquiry is economic science, specifically, the attention will converge upon the scheme of material life and will take in other phases of civilization only in their correlation with the scheme of material civilization.

Like all human culture this material civilization is a scheme of institutions – institutional fabric and institutional growth. But institutions are an outgrowth of habit. The growth of culture is a cumulative sequence of habituation, and the ways and means of it are the habitual response of human nature to exigencies that vary incontinently, cumulatively, but with something of a consistent sequence in the cumulative variations that so go forward, – incontinently, because each new move creates a new situation which induces a further new variation in the habitual manner of response; cumulatively, because each new situation is a variation of what has gone before it and embodies as causal factors all that has been effected by what went before; consistently, because the underlying traits of human nature (propensities, aptitudes, and what not) by force of which the response takes place, and on the ground of which the habituation takes effect, remain substantially unchanged.

Evidently an economic inquiry which occupies itself exclusively with the movements of this consistent, elemental human nature under given, stable institutional conditions – such as is the case with the current hedonistic economics – can reach statical results alone; since it makes abstraction from those elements that make for anything but a statical result. On the other hand an adequate theory of economic conduct, even for statical purposes, cannot be drawn in terms of the individual simply – as is the case in the marginal-utility economics – because it cannot be drawn in terms of the underlying traits of human nature simply; since the response that goes to make up human conduct takes place under institutional norms and only under stimuli that have an institutional bearing; for the situation that provokes and inhibits action in any given case is itself in great part of institutional, cultural derivation. Then, too, the phenomena of human life occur only as phenomena of the life of a group or community: only under stimuli due to contact with the group and only under the (habitual) control exercised by canons of conduct imposed by the group's scheme of life. Not only is the individual's conduct hedged about and directed by his habitual relations to his fellows in the group, but these relations, being of an institutional character, vary as the institutional scheme varies. The wants and desires, the end and aim, the ways and means, the amplitude and drift of the individual's conduct are functions of an institutional variable that is of a highly complex and wholly unstable character.

The growth and mutations of the institutional fabric are an outcome of the conduct of the individual members of the group, since it is out of the experience of the individuals, through the habituation of individuals, that institutions arise; and it is in this same experience that these institutions act to direct and define the aims and end of conduct. It is, of course, on individuals that the system of institutions imposes those conventional standards, ideals, and canons of conduct that make up the community's scheme of life. Scientific inquiry in this field, therefore, must deal with individual conduct and must formulate its theoretical results in terms of individual conduct. But such an inquiry can serve the purposes of a genetic theory only if and in so far as this individual conduct is attended to in those respects in which it counts toward habituation, and so toward change (or stability) of the institutional fabric, on the one hand, and in those respects in which it is prompted and guided by the received institutional conceptions and ideals on the other hand. The postulates of marginal utility, and the hedonistic preconceptions generally, fail at this point in that they confine the attention to such bearings of economic conduct as are conceived not to be conditioned by habitual standards and ideals and to have no effect in the way of habituation. They disregard or abstract from the causal sequence of propensity and habituation in economic life and exclude from theoretical inquiry all such interest in the facts of cultural growth, in order to attend to those features of the case that are conceived to be idle in this respect. All such facts of institutional force and growth are put on one side as not being germane to pure theory; they are to be taken account of, if at all, by afterthought, by a more or less vague and general allowance for inconsequential disturbances due to occasional human infirmity. Certain institutional phenomena, it is true, are comprised among the premises of the hedonists, as has been noted above; but they are included as postulates a priori. So the institution of ownership is taken into the inquiry not as a factor of growth or an element subject to change, but as one of the primordial and immutable facts of the order of nature, underlying the hedonistic calculus. Property, ownership, is presumed as the basis of hedonistic discrimination and it is conceived to be given in its finished (nineteenth-century) scope and force. There is no thought either of a conceivable growth of this definitive nineteenth-century institution out of a cruder past or of any conceivable cumulative change in the scope and force of ownership in the present or future. Nor is it conceived that the presence of this institutional element in men's economic relations in any degree affects or disguises the hedonistic calculus, or that its pecuniary conceptions and standards in any degree standardize, color, mitigate, or divert the hedonistic calculator from the direct and unhampered quest of the net sensuous gain. While the institution of property is included in this way among the postulates of the theory, and is even presumed to be ever-present in the economic situation, it is allowed to have no force in shaping economic conduct, which is conceived to run its course to its hedonistic outcome as if no such institutional factor intervened between the impulse and its realization. The institution of property, together with all the range of pecuniary conceptions that belong under it and that cluster about it, are presumed to give rise to no habitual or conventional canons of conduct or standards of valuation, no proximate ends, ideals, or aspirations. All pecuniary notions arising from ownership are treated simply as expedients of computation which mediate between the pain-cost and the pleasure-gain of hedonistic choice, without lag, leak, or friction; they are conceived simply as the immutably correct, God-given notation of the hedonistic calculus.

The modern economic situation is a business situation, in that economic activity of all kinds is commonly controlled by business considerations. The exigencies of modern life are commonly pecuniary exigencies. That is to say they are exigencies of the ownership of property. Productive efficiency and distributive gain are both rated in terms of price. Business considerations are considerations of price, and pecuniary exigencies of whatever kind in the modern communities are exigencies of price. The current economic situation is a price system. Economic institutions in the modern civilized scheme of life are (prevailingly) institutions of the price system. The accountancy to which all phenomena of modern economic life are amenable is an accountancy in terms of price; and by the current convention there is no other recognized scheme of accountancy, no other rating, either in law or in fact, to which the facts of modern life are held amenable. Indeed, so great and pervading a force has this habit (institution) of pecuniary accountancy become that it extends, often as a matter of course, to many facts which properly have no pecuniary bearing and no pecuniary magnitude, as, *e.g.*, works of art, science, scholarship, and religion. More or less freely and fully, the price system dominates the current commonsense in its appreciation and rating of these non-pecuniary ramifications of modern culture; and this in spite of the fact that, on reflection, all men of normal intelligence will freely admit that these matters lie outside the scope of pecuniary valuation.

Current popular taste and the popular sense of merit and demerit are notoriously affected in some degree by pecuniary considerations. It is a matter of common notoriety, not to be denied or explained away, that pecuniary ("commercial") tests and standards are habitually made use of outside of commercial interests proper. Precious stones, it is admitted, even by hedonistic economists, are more esteemed than they would be if they were more plentiful and cheaper. A wealthy person meets with more consideration and enjoys a larger measure of good repute than would fall to the share of the same person with the same habit of mind and body and the same record of good and evil deeds if he were poorer. It may well be that this current "commercialisation" of taste and appreciation has been overstated by superficial and hasty critics of contemporary life, but it will not be denied that there is a modicum of truth in the allegation. Whatever substance it has, much or little, is due to carrying over into other fields of interest the habitual conceptions induced by dealing with and thinking of pecuniary matters. These "commercial" conceptions of merit and demerit are derived from business experience. The pecuniary tests and standards so applied outside of business transactions and relations are not reducible to sensuous terms of pleasure and pain. Indeed, it may, *e.g.*, be true, as is commonly believed, that the contemplation of a wealthy neighbor's pecuniary superiority yields painful rather than pleasurable sensations as an immediate result; but it is equally true that such a wealthy neighbor is, on the whole, more highly regarded and more considerately treated than another neighbor who differs from the former only in being less enviable in respect of wealth.

It is the institution of property that gives rise to these habitual grounds of discrimination, and in modern times, when wealth is counted in terms of money, it is in terms of money value that these tests and standards of pecuniary excellence are applied. This much will be admitted. Pecuniary institutions induce pecuniary habits of thought which affect men's discrimination outside of pecuniary matters; but the hedonistic interpretation alleges that such pecuniary habits of thought do not affect men's discrimination in pecuniary matters. Although the institutional scheme of the price system visibly dominates the modern community's thinking in matters that lie outside the economic interest, the hedonistic economists insist, in effect, that this institutional scheme must be accounted of no effect within that range of activity to which it owes its genesis, growth, and persistence. The phenomena of business, which are peculiarly and uniformly phenomena of price, are in the scheme of the hedonistic theory reduced to non-pecuniary hedonistic terms and the theoretical formulation is carried out as if pecuniary conceptions had no force within the traffic in which such conceptions originate. It is admitted that preoccupation with commercial interests has "commercialised" the rest of modern life, but the "commercialisation" of commerce is not admitted. Business transactions and computations in pecuniary terms, such as loans, discounts, and capitalisation, are without hesitation or abatement converted into terms of hedonistic utility, and conversely.

It may be needless to take exception to such conversion from pecuniary into sensuous terms, for the theoretical purpose for which it is habitually made; although, if need were, it might not be excessively difficult to show that the whole hedonistic basis of such a conversion is a psychological misconception. But it is to the remoter theoretical consequences of such a conversion that exception is to be taken. In making the conversion abstraction is made from whatever elements do not lend themselves to its terms; which amounts to abstracting from precisely those elements of business that have an institutional force and that therefore would lend themselves to scientific inquiry of the modern kind – those (institutional) elements whose analysis might contribute to an understanding of modern business and of the life of the modern business community as contrasted with the assumed primordial hedonistic calculus.

The point may perhaps be made clearer. Money and the habitual resort to its use are conceived to be simply the ways and means by which consumable goods are acquired, and therefore simply a convenient method by which to procure the pleasurable sensations of consumption; these latter being in hedonistic theory the sole and overt end of all economic endeavor. Money values have therefore no other significance than that of purchasing power over consumable goods, and money is simply an expedient of computation. Investment, credit extensions, loans of all kinds and degrees, with payment of interest and the rest, are likewise taken simply as intermediate steps between the pleasurable sensations of consumption and the efforts induced by the anticipation of these sensations, other bearings of the case being disregarded. The balance being kept in terms of the hedonistic consumption, no disturbance arises in this pecuniary traffic so long as the extreme terms of this extended hedonistic equation – pain-cost and pleasure-gain – are not altered, what lies between these extreme terms being merely algebraic notation employed for convenience of accountancy. But such is not the run of the facts in modern business. Variations of capitalization, *e.g.*, occur without its being practicable to refer them to visibly equivalent variations either in the state of the industrial arts or in the sensations of consumption. Credit extensions tend to inflation of credit, rising prices, overstocking of markets, etc., likewise without a visible or securely traceable correlation in the state of the industrial arts or in the pleasures of consumption; that is to say, without a visible basis in those material elements to which the hedonistic theory reduces all economic phenomena. Hence the run of the facts, in so far, must be thrown out of the theoretical formulation. The hedonistically presumed final purchase of consumable goods is habitually not contemplated in the pursuit of business enterprise. Business men habitually aspire to accumulate wealth in excess of the limits of practicable consumption, and the wealth so accumulated is not intended to be converted by a final transaction of purchase into consumable goods or sensations of consumption. Such commonplace facts as these, together with the endless web of business detail of a like pecuniary character, do not in hedonistic theory raise a question as to how these conventional aims, ideals, aspirations, and standards have come into force or how they affect the scheme of life in business or outside of it; they do not raise those questions because such questions cannot be answered in the terms which the hedonistic economists are content to use, or, indeed, which their premises permit them to use. The question which arises is how to explain the facts away: how theoretically to neutralize them so that they will not have to appear in the theory, which can then be drawn in direct and unambiguous terms of rational hedonistic calculation. They are explained away as being aberrations due to oversight or lapse of memory on the part of business men, or to some failure of logic or insight. Or they are construed and interpreted into the rationalistic terms of the hedonistic calculus by resort to an ambiguous use of the hedonistic concepts. So that the whole "money economy," with all the machinery of credit and the rest, disappears in a tissue of metaphors to reappear theoretically expurgated, sterilized, and simplified into a "refined system of barter," culminating in a net aggregate maximum of pleasurable sensations of consumption.

But since it is in just this unhedonistic, unrationalistic pecuniary traffic that the tissue of business life consists; since it is this peculiar conventionalism of aims and standards that differentiates the life of the modern business community from any conceivable earlier or cruder phase of economic life; since it is in this tissue of pecuniary intercourse and pecuniary concepts, ideals, expedients, and aspirations that the conjunctures of business life arise and run their course of felicity and devastation; since it is here that those institutional changes take place which distinguish one phase or era of the business community's life from any other; since the growth and change of these habitual, conventional elements make the growth and character of any business era or business community; any theory of business which sets these elements aside or explains them away misses the main facts which it has gone out to seek. Life and its conjunctures and institutions being of this complexion, however much that state of the case may be deprecated, a theoretical account of the phenomena of this life must be drawn in these terms in which the phenomena occur. It is not simply that the hedonistic interpretation of modern economic phenomena is inadequate or misleading; if the phenomena are subjected to the hedonistic interpretation in the theoretical analysis they disappear from the theory; and if they would bear the interpretation in fact they would disappear in fact. If, in fact, all the conventional relations and principles of pecuniary intercourse were subject to such a perpetual rationalized, calculating revision, so that each article of usage, appreciation, or procedure must approve itself *de novo* on hedonistic grounds of sensuous expediency to all concerned at every move, it is not conceivable that the institutional fabric would last over night.

Note

1. The conduct of mankind differs from that of the brutes in being determined by anticipated sensations of pleasure and pain, instead of actual sensations. Hereby, in so far, human conduct is taken out of the sequence of cause and effect and falls instead under the rule of sufficient reason. By virtue of this rational faculty in man the connection between stimulus and response is teleological instead of causal.

   The reason for assigning the first and decisive place to pleasure, rather than to pain, in the determination of human conduct, appears to be the (tacit) acceptance of that optimistic doctrine of a beneficent order of nature which the nineteenth century inherited from the eighteenth.

PART TWO

POSITIVIST AND POPPERIAN VIEWS

The development of logical positivism and of Karl Popper's views (see the introduction to this volume) had a significant impact on the methodology of economics. Economists such as Terence Hutchison and Paul Samuelson noted that much of economic theory appeared not to satisfy logical positivist or Popperian standards of theory assessment, and the 1930s and 1940s saw naive tests of fundamental principles of the theory of the firm, which appeared to refute them. This serious challenge was met mainly by Milton Friedman, whose essay reprinted here is the most influential methodological tract of modern times. It has been subjected to a barrage of criticism, of which Herbert Simon's and my own brief comments are only a tiny sample.

Although Imre Lakatos's work on philosophy of science dates from much later and shows the influence of Thomas Kuhn's views, Lakatos's views on theory assessment are closely related to Popper's, and both were influential among economic methodologists in the 1970s and 1980s. The last essay in this part by D. Wade Hands compactly summarizes the issues that arise in applying Karl Popper's and Imre Lakatos's views to economics.

SEVEN

The Methodology of Positive Economics

Milton Friedman

Milton Friedman (1912–2006) was born in Brooklyn, New York, and received his Ph.D. in economics from Columbia University. He taught at the University of Minnesota, and then for many years at the University of Chicago. After 1977, he was a Senior Research Fellow at the Hoover Institution in Stanford, California. Friedman is best known for his work in monetary theory and for his concern for free enterprise and individual liberty. Milton Friedman was awarded the Nobel Prize in economics in 1976. The following essay, which is reprinted in its entirety, is the most influential work on economic methodology of this century.

In his admirable book on *The Scope and Method of Political Economy* John Neville Keynes distinguishes among "a *positive science* . . . [,] a body of systematized knowledge concerning what is; a *normative* or *regulative science* . . . [,] a body of systematized knowledge discussing criteria of what ought to be . . . ; an *art* . . . [,] a system of rules for the attainment of a given end"; comments that "confusion between them is common and has been the source of many mischievous errors"; and urges the importance of "recognizing a distinct positive science of political economy."[1]

This [essay] is concerned primarily with certain methodological problems that arise in constructing the "distinct positive science" Keynes called for – in particular, the problem how to decide whether a suggested hypothesis or theory should be tentatively accepted as part of the "body of systematized

---

I have incorporated bodily in this article without special reference most of my brief "Comment" in *A Survey of Contemporary Economics*, Vol. II (B. F. Haley, ed.) (Chicago: Richard D. Irwin, Inc., 1952), pp. 455–7.

I am indebted to Dorothy S. Brady, Arthur F. Burns, and George J. Stigler for helpful comments and criticism.

knowledge concerning what is." But the confusion Keynes laments is still so rife and so much of a hindrance to the recognition that economics can be, and in part is, a positive science that it seems well to preface the main body of the paper with a few remarks about the relation between positive and normative economics.

I. The Relation between Positive and Normative Economics

Confusion between positive and normative economics is to some extent inevitable. The subject matter of economics is regarded by almost everyone as vitally important to himself and within the range of his own experience and competence; it is the source of continuous and extensive controversy and the occasion for frequent legislation. Self-proclaimed "experts" speak with many voices and can hardly all be regarded as disinterested; in any event, on questions that matter so much, "expert" opinion could hardly be accepted solely on faith even if the "experts" were nearly unanimous and clearly disinterested.[2] The conclusions of positive economics seem to be, and are, immediately relevant to important normative problems, to questions of what ought to be done and how any given goal can be attained. Laymen and experts alike are inevitably tempted to shape positive conclusions to fit strongly held normative preconceptions and to reject positive conclusions if their normative implications – or what are said to be their normative implications – are unpalatable.

Positive economics is in principle independent of any particular ethical position or normative judgments. As Keynes says, it deals with "what is," not with "what ought to be." Its task is to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances. Its performance is to be judged by the precision, scope, and conformity with experience of the predictions it yields. In short, positive economics is, or can be, an "objective" science, in precisely the same sense as any of the physical sciences. Of course, the fact that economics deals with the interrelations of human beings, and that the investigator is himself part of the subject matter being investigated in a more intimate sense than in the physical sciences, raises special difficulties in achieving objectivity at the same time that it provides the social scientist with a class of data not available to the physical scientist. But neither the one nor the other is, in my view, a fundamental distinction between the two groups of sciences.[3]

Normative economics and the art of economics, on the other hand, cannot be independent of positive economics. Any policy conclusion necessarily rests on a prediction about the consequences of doing one thing rather than another, a prediction that must be based – implicitly or explicitly – on positive economics. There is not, of course, a one-to-one relation between policy conclusions and the conclusions of positive economics; if there were, there would be no separate normative science. Two individuals may agree on the consequences of a particular piece of legislation. One may regard them as desirable on balance and so favor the legislation; the other, as undesirable and so oppose the legislation.

I venture the judgment, however, that currently in the Western world, and especially in the United States, differences about economic policy among disinterested citizens derive predominantly from different predictions about the economic consequences of taking action – differences that in principle can be eliminated by the progress of positive economics – rather than from fundamental differences in basic values, differences about which men can ultimately only fight. An obvious and not unimportant example is minimum-wage legislation. Underneath the welter of arguments offered for and against such legislation there is an underlying consensus on the objective of achieving a "living wage" for all, to use the ambiguous phrase so common in such discussions. The difference of opinion is largely grounded on an implicit or explicit difference in predictions about the efficacy of this particular means in furthering the agreed-on end. Proponents believe (predict) that legal minimum wages diminish poverty by raising the wages of those receiving less than the minimum wage as well as of some receiving more than the minimum wage without any counterbalancing increase in the number of people entirely unemployed or employed less advantageously than they otherwise would be. Opponents believe (predict) that legal minimum wages increase poverty by increasing the number of people who are unemployed or employed less advantageously and that this more than offsets any favorable effect on the wages of those who remain employed. Agreement about the economic consequences of the legislation might not produce complete agreement about its desirability, for differences might still remain about its political or social consequences; but, given agreement on objectives, it would certainly go a long way toward producing consensus.

Closely related differences in positive analysis underlie divergent views about the appropriate role and place of trade-unions and the desirability of direct price and wage controls and of tariffs. Different predictions about the importance of so-called "economics of scale" account very largely for divergent views about the desirability or necessity of detailed government regulation of industry and even of socialism rather than private enterprise. And this list could be extended indefinitely.[4] Of course, my judgment that the major differences about economic policy in the Western world are of this kind is itself a "positive" statement to be accepted or rejected on the basis of empirical evidence.

If this judgment is valid, it means that a consensus on "correct" economic policy depends much less on the progress of normative economics proper than on the progress of a positive economics yielding conclusions that are, and deserve to be, widely accepted. It means also that a major reason for distinguishing positive economics sharply from normative economics is precisely the contribution that can thereby be made to agreement about policy.

II. Positive Economics

The ultimate goal of a positive science is the development of a "theory" or "hypothesis" that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed. Such a theory is, in general, a complex intermixture of two elements. In part, it is a "language" designed to promote "systematic and organized methods of reasoning."[5] In part, it is a body of substantive hypotheses designed to abstract essential features of complex reality.

Viewed as a language, theory has no substantive content; it is a set of tautologies. Its function is to serve as a filing system for organizing empirical material and facilitating our understanding of it; and the criteria by which it is to be judged are those appropriate to a filing system. Are the categories clearly and precisely defined? Are they exhaustive? Do we know where to file each individual item, or is there considerable ambiguity? Is the system of headings and subheadings so designed that we can quickly find an item we want, or must we hunt from place to place? Are the items we shall want to consider jointly filed together? Does the filing system avoid elaborate cross-references?

The answers to these questions depend partly on logical, partly on factual, considerations. The canons of formal logic alone can show whether a particular language is complete and consistent, that is, whether propositions in the language are "right" or "wrong." Factual evidence alone can show whether the categories of the "analytical filing system" have a meaningful empirical counterpart, that is, whether they are useful in analyzing a particular class of concrete problems.[6] The simple example of "supply" and "demand" illustrates both this point and the preceding list of analogical questions. Viewed as elements of the language of economic theory, these are the two major categories into which factors affecting the relative prices of products or factors of production are classified. The usefulness of the dichotomy depends on the "empirical generalization that an enumeration of the forces affecting demand in any problem and of the forces affecting supply will yield two lists that contain few items in common."[7] Now this generalization is valid for markets like the final market for a consumer good. In such a market there is a clear and sharp distinction between the economic units that can be regarded as demanding the product and those that can be regarded as supplying it. There is seldom much doubt whether a particular factor should be classified as affecting supply, on the one hand, or demand, on the other; and there is seldom much necessity for considering cross-effects (cross-references) between the two categories. In these cases the simple and even obvious step of filing the relevant factors under the headings of "supply" and "demand" effects a great simplification of the problem and is an effective safeguard against fallacies that otherwise tend to occur. But the generalization is not always valid. For example, it is not valid for the day-to-day fluctuations of prices in a primarily speculative market. Is a rumor of an increased excess-profits tax, for example, to be regarded as a factor operating primarily on today's supply of corporate equities in the stock market or on today's demand for them? In similar fashion, almost every factor can with about as much justification be classified under the heading "supply" as under the heading "demand." These concepts can still be used and may not be entirely pointless; they are still "right" but clearly less useful than in the first example because they have no meaningful empirical counterpart.

Viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena which it is intended to "explain." Only factual evidence can show whether it is "right" or "wrong" or, better, tentatively "accepted" as valid or "rejected." As I shall argue at greater length below, the only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience. The hypothesis is rejected if its predictions are contradicted ("frequently" or more often than predictions from an alternative hypothesis); it is accepted if its predictions are not contradicted; great confidence is attached to it if it has survived many opportunities for contradiction. Factual evidence can never "prove" a hypothesis; it can only fail to disprove it, which is what we generally mean when we say, somewhat inexactly, that the hypothesis has been "confirmed" by experience.

To avoid confusion, it should perhaps be noted explicitly that the "predictions" by which the validity of a hypothesis is tested need not be about phenomena that have not yet occurred, that is, need not be forecasts of future events; they may be about phenomena that have occurred but observations on which have not yet been made or are not known to the person making the prediction. For example, a hypothesis may imply that such and such must have happened in 1906, given some other known circumstances. If a search of the records reveals that such and such did happen, the prediction is confirmed; if it reveals that such and such did not happen, the prediction is contradicted.

The validity of a hypothesis in this sense is not by itself a sufficient criterion for choosing among alternative hypotheses. Observed facts are necessarily finite in number; possible hypotheses, infinite. If there is one hypothesis that is consistent with the available evidence, there are always an infinite number that are.[8] For example, suppose a specific excise tax on a particular commodity produces a rise in price equal to the amount of the tax. This is consistent with competitive conditions, a stable demand curve, and a horizontal and stable supply curve. But it is also consistent with competitive conditions and a positively or negatively sloping supply curve with the required compensating shift in the demand curve or the supply curve; with monopolistic conditions, constant marginal costs, and stable demand curve, of the particular shape required to produce this result; and so on indefinitely. Additional evidence with which the hypothesis is to be consistent may rule out some of these possibilities; it can never reduce them to a single possibility alone capable of being consistent with the finite evidence. The choice among alternative hypotheses equally consistent with the available evidence must to some extent be arbitrary, though there is general agreement that relevant considerations are suggested by the criteria "simplicity" and "fruitfulness," themselves notions that defy completely objective specification. A theory is "simpler" the less the initial knowledge needed to make a prediction within a given field of phenomena; it is more "fruitful" the more precise the resulting prediction, the wider the area within which the theory yields predictions, and the more additional lines for further research it suggests. Logical completeness and consistency are relevant but play a subsidiary role; their function is to assure that the hypothesis says what it is intended to say and does so alike for all users – they play the same role here as checks for arithmetical accuracy do in statistical computations.

Unfortunately, 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. Generally, we must rely on evidence cast up by the "experiments" that happen to occur. The inability to conduct so-called "controlled experiments" does not, in my view, reflect a basic difference between the social and physical sciences both because it is not peculiar to the social sciences – witness astronomy – and because the distinction between a controlled experiment and uncontrolled experience is at best one of degree. No experiment can be completely controlled, and every experience is partly controlled, in the sense that some disturbing influences are relatively constant in the course of it.

Evidence cast up by experience is abundant and frequently as conclusive as that from contrived experiments; thus the inability to conduct experiments is not a fundamental obstacle to testing hypotheses by the success of their predictions. But such evidence is far more difficult to interpret. It is frequently complex and always indirect and incomplete. Its collection is often arduous, and its interpretation generally requires subtle analysis and involved chains of reasoning, which seldom carry real conviction. The denial to economics of the dramatic and direct evidence of the "crucial" experiment does hinder the adequate testing of hypotheses; but this is much less significant than the difficulty it places in the way of achieving a reasonably prompt and wide consensus on the conclusions justified by the available evidence. It renders the weeding-out of unsuccessful hypotheses slow and difficult. They are seldom downed for good and are always cropping up again.

There is, of course, considerable variation in these respects. Occasionally, experience casts up evidence that is about as direct, dramatic, and convincing as any that could be provided by controlled experiments. Perhaps the most obviously important example is the evidence from inflations on the hypothesis that a substantial increase in the quantity of money within a relatively short period is accompanied by a substantial increase in prices. Here the evidence is dramatic, and the chain of reasoning required to interpret it is relatively short. Yet, despite numerous instances of substantial rises in prices, their essentially one-to-one correspondence with substantial rises in the stock of money, and the wide variation in other circumstances that might appear to be relevant, each new experience of inflation brings forth vigorous contentions, and not only by the lay public, that the rise in the stock of money is either an incidental effect of a rise in prices produced by other factors or a purely fortuitous and unnecessary concomitant of the price rise.

One effect of the difficulty of testing substantive economic hypotheses has been to foster a retreat into purely formal or tautological analysis.[9] As already noted, tautologies have an extremely important place in economics and other sciences as a specialized language or "analytical filing system." Beyond this, formal logic and mathematics, which are both tautologies, are essential aids in checking the correctness of reasoning, discovering the implications of hypotheses, and determining whether supposedly different hypotheses may not really be equivalent or wherein the differences lie.

But economic theory must be more than a structure of tautologies if it is to be able to predict and not merely describe the consequences of action; if it is to be something different from disguised mathematics.[10] And the usefulness of the tautologies themselves ultimately depends, as noted above, on the acceptability of the substantive hypotheses that suggest the particular categories into which they organize the refractory empirical phenomena.

A more serious effect of the difficulty of testing economic hypotheses by their predictions is to foster misunderstanding of the role of empirical evidence in theoretical work. Empirical evidence is vital at two different, though closely related, stages: in constructing hypotheses and in testing their validity. Full and comprehensive evidence on the phenomena to be generalized or "explained" by a hypothesis, besides its obvious value in suggesting new hypotheses, is needed to assure that a hypothesis explains what it sets out to explain – that its implications for such phenomena are not contradicted in advance by experience that has already been observed.[11] Given that the hypothesis is consistent with the evidence at hand, its further testing involves deducing from it new facts capable of being observed but not previously known and checking these deduced facts against additional empirical evidence. For this test to be relevant, the deduced facts must be about the class of phenomena the hypothesis is designed to explain; and they must be well enough defined so that observation can show them to be wrong.

The two stages of constructing hypotheses and testing their validity are related in two different respects. In the first place, the particular facts that enter at each stage are partly an accident of the collection of data and the knowledge of the particular investigator. The facts that serve as a test of the implications of a hypothesis might equally well have been among the raw material used to construct it, and conversely. In the second place, the process never begins from scratch; the so-called "initial stage" itself always involves comparison of the implications of an earlier set of hypotheses with observation; the contradiction of these implications is the stimulus to the construction of new hypotheses or revision of old ones. So the two methodologically distinct stages are always proceeding jointly.

Misunderstanding about this apparently straightforward process centers on the phrase "the class of phenomena the hypothesis is designed to explain." The difficulty in the social sciences of getting new evidence for this class of phenomena and of judging its conformity with the implications of the hypothesis makes it tempting to suppose that other, more readily available, evidence is equally relevant to the validity of the hypothesis – to suppose that hypotheses have not only "implications" but also "assumptions" and that the conformity of these "assumptions" to "reality" is a test of the validity of the hypothesis *different from* or *additional to* the test by implications, this widely held view is fundamentally wrong and productive of much mischief. Far from providing an easier means for sifting valid from invalid hypotheses, it only confuses the issue, promotes misunderstanding about the significance of empirical evidence for economic theory, produces a misdirection of much intellectual effort devoted to the development of positive economics, and impedes the attainment of consensus on tentative hypotheses in positive economics.

In so far as a theory can be said to have "assumptions" at all, and in so far as their "realism" can be judged independently of the validity of predictions, the relation between the significance of a theory and the "realism" of its "assumptions" is almost the opposite of that suggested by the view under criticism. Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense).[12] The reason is simple. A hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. To be important, therefore, a hypothesis must be descriptively false in its assumptions; it takes account of, and accounts for, none of the many other attendant circumstances, since its very success shows them to be irrelevant for the phenomena to be explained.

To put this point less paradoxically, the relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic," for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions. The two supposedly independent tests thus reduce to one test.

The theory of monopolistic and imperfect competition is one example of the neglect in economic theory of these propositions. The development of this analysis was explicitly motivated, and its wide acceptance and approval largely explained, by the belief that the assumptions of "perfect competition" or "perfect monopoly" said to underlie neoclassical economic theory are a false image of reality. And this belief was itself based almost entirely on the directly perceived descriptive inaccuracy of the assumptions rather than on any recognized contradiction of predictions derived from neoclassical economic theory. The lengthy discussion on marginal analysis in the

American Economic Review some years ago is an even clearer, though much less important, example. The articles on both sides of the controversy largely neglect what seems to me clearly the main issue – the conformity to experience of the implications of the marginal analysis – and concentrate on the largely irrelevant question whether businessmen do or do not in fact reach their decisions by consulting schedules, or curves, or multivariable functions showing marginal cost and marginal revenue.[13] Perhaps these two examples, and the many others they readily suggest, will serve to justify a more extensive discussion of the methodological principles involved than might otherwise seem appropriate.

III. Can a Hypothesis be Tested by the Realism of its Assumptions?

We may start with a simple physical example, the law of falling bodies. It is an accepted hypothesis that the acceleration of a body dropped in a vacuum is a constant – g, or approximately 32 feet per second per second on the earth – and is independent of the shape of the body, the manner of dropping it, etc. This implies that the distance traveled by a falling body in any specified time is given by the formula $s = 1/2\ gt^2$, where s is the distance traveled in feet and t is time in seconds. The application of this formula to a compact ball dropped from the roof of a building is equivalent to saying that a ball so dropped behaves *as if* it were falling in a vacuum. Testing this hypothesis by its assumptions presumably means measuring the actual air pressure and deciding whether it is close enough to zero. At sea level the air pressure is about 15 pounds per square inch. Is 15 sufficiently close to zero for the difference to be judged insignificant? Apparently it is, since the actual time taken by a compact ball to fall from the roof of a building to the ground is very close to the time given by the formula. Suppose, however, that a feather is dropped instead of a compact ball. The formula then gives wildly inaccurate results. Apparently, 15 pounds per square inch is significantly different from zero for a feather but not for a ball. Or, again, suppose the formula is applied to a ball dropped from an airplane at an altitude of 30,000 feet. The air pressure at this altitude is decidedly less than 15 pounds per square inch. Yet, the actual time of fall from 30,000 feet to 20,000 feet, at which point the air pressure is still much less than at sea level, will differ noticeably from the time predicted by the formula – much more noticeably than the time taken by a compact ball to fall from the roof of a building to the ground. According to the formula, the velocity of the ball should be gt and should therefore increase steadily. In fact, a ball dropped at 30,000 feet will reach its top velocity well before it hits the ground. And similarly with other implications of the formula.

The initial question whether 15 is sufficiently close to zero for the difference to be judged insignificant is clearly a foolish question by itself. Fifteen pounds per square inch is 2,160 pounds per square foot, or 0.0075 ton per square inch. There is no possible basis for calling these numbers "small" or "large" without some external standard of comparison. And the only relevant standard of comparison is the air pressure for which the formula does or does not work under a given set of circumstances. But this raises the same problem at a second level. What is the meaning of "does or does not work"? Even if we could eliminate errors of measurement, the measured time of fall would seldom if ever be precisely equal to the computed time of fall. How large must the difference between the two be to justify saying that the theory "does not work"? Here there are two important external standards of comparison. One is the accuracy achievable by an alternative theory with which this theory is being compared and which is equally acceptable on all other grounds. The other arises when there exists a theory that is known to yield better predictions but only at a greater cost. The gains from greater accuracy, which depend on the purpose in mind, must then be balanced against the costs of achieving it.

The example illustrates both the impossibility of testing a theory by its assumptions and also the ambiguity of the concept "the assumptions of a theory." The formula $s = 1/2\ gt^2$ is valid for bodies falling in a vacuum and can be derived by analyzing the behavior of such bodies. It can therefore be stated: under a wide range of circumstances, bodies that fall in the actual atmosphere behave *as if* they were falling in a vacuum. In the language so common in economics this would be rapidly translated into: the formula assumes a vacuum. Yet it clearly does no such thing. What it does say is that in many cases the existence of air pressure, the shape of the body, the name of the person dropping the body, the kind of mechanism used to drop the body, and a host of other attendant circumstances have no appreciable effect on the distance the body falls in a specified time. The hypothesis can readily be rephrased to omit all mention of a vacuum: under a wide range of circumstances, the distance a body falls in a specified time is given by the formula $s = 1/2\ gt^2$. The history of this formula and its associated physical theory aside, is it meaningful to say that it assumes a vacuum? For all I know there may be other sets of assumptions that would yield the same formula. The formula is accepted because it works, not because we live in an approximate vacuum – whatever that means.

The important problem in connection with the hypothesis is to specify the circumstances under which the formula works or, more precisely, the general magnitude of the error in its predictions under various circumstances. Indeed, as is implicit in the above rephrasing of the hypothesis, such a specification is not one thing and the hypothesis another. The specification is itself an essential part of the hypothesis, and it is a part that is peculiarly likely to be revised and extended as experience accumulates.

In the particular case of falling bodies a more general, though still incomplete, theory is available, largely as a result of attempts to explain the errors of the simple theory, from which the influence of some of the possible disturbing factors can be calculated and of which the simple theory is a special case. However, it does not always pay to use the more general theory because the extra accuracy it yields may not justify the extra cost of using it, so the question under what circumstances the simpler theory works "well enough" remains important. Air pressure is one, but only one, of the variables that define these circumstances; the shape of the body, the velocity attained, and still other variables are relevant as well. One way of interpreting the variables other than air pressure is to regard them as determining whether a particular departure from the "assumption" of a vacuum is or is not significant. For example, the difference in shape of the body can be said to make 15 pounds per square inch significantly different from zero for a feather but not for a compact ball dropped a moderate distance. Such a statement must, however, be sharply distinguished from the very different statement that the theory does not work for a feather because its assumptions are false. The relevant relation runs the other way: the assumptions are false for a feather because the theory does not work. This point needs emphasis, because the entirely valid use of "assumptions" in *specifying* the circumstances for which theory holds is frequently, and erroneously, interpreted to mean that the assumptions can be used to *determine* the circumstances for which a theory holds, and has, in this way, been an important source of the belief that a theory can be tested by its assumptions.

Let us turn now to another example, this time a constructed one designed to be an analogue of many hypotheses in the social sciences. Consider the density of leaves around a tree. I suggest the hypothesis that the leaves are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives, given the position of its neighbors, as if it knew the physical laws determining the amount of sunlight that would be received in various positions and could move rapidly or instantaneously from any one position to any other desired and unoccupied position.[14] Now some of the more obvious implications of this hypothesis are clearly consistent with experience: for example, leaves are in general denser on the south than on the north side of trees but, as the hypothesis implies, less so or not at all on the northern slope of a hill or when the south side of the trees is shaded in some other way. Is the hypothesis rendered unacceptable or invalid because, so far as we know, leaves do not "deliberate" or consciously "seek," have not been to school and learned the relevant laws of science or the mathematics required to calculate the "optimum" position, and cannot move from position to position? Clearly, none of these contradictions of the hypothesis is vitally relevant; the phenomena involved are not within the "class of phenomena the hypothesis is designed to explain"; the hypothesis does not assert that leaves do these things but only that their density is the same *as if* they did. Despite the apparent falsity of the "assumptions" of the hypothesis, it has great plausibility because of the conformity of its implications with observation. We are inclined to "explain" its validity on the ground that sunlight contributes to the growth of leaves and that hence leaves will grow denser or more putative leaves survive where there is more sun, so the result achieved by purely passive adaptation to external circumstances is the same as the result that would be achieved by deliberate accommodation to them. This alternative hypothesis is more attractive than the constructed hypothesis not because its "assumptions" are more "realistic" but rather because it is part of a more general theory that applies to a wider variety of phenomena, of which the position of leaves around a tree is a special case, has more implications capable of being contradicted, and has failed to be contradicted under a wider variety of circumstances. The direct evidence for the growth of leaves is in this way strengthened by the indirect evidence from the other phenomena to which the more general theory applies.

The constructed hypothesis is presumably valid, that is, yields "sufficiently" accurate predictions about the density of leaves, only for a particular class of circumstances. I do not know what these circumstances are or how to define them. It seems obvious, however, that in this example the "assumptions" of the theory will play no part in specifying them: the kind of tree, the character of the soil, etc., are the types of variables that are likely to define its range of validity, not the ability of the leaves to do complicated mathematics or to move from place to place.

A largely parallel example involving human behavior has been used elsewhere by Savage and me.[15] Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc., describing the location of the balls, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players.

It is only a short step from these examples to the economic hypothesis that under a wide range of circumstances individual firms behave *as if* they were seeking rationally to maximize their expected returns (generally if misleadingly called "profits")[16] and had full knowledge of the data needed to succeed in this attempt; *as if*, that is, they knew the relevant cost and demand functions, calculated marginal cost and marginal revenue from all actions open to them, and pushed each line of action to the point at which the relevant marginal cost and marginal revenue were equal. Now, of course, businessmen do not actually and literally solve the system of simultaneous equations in terms of which the mathematical economist finds it convenient to express this hypothesis, any more than leaves or billiard players explicitly go through complicated mathematical calculations or falling bodies decide to create a vacuum. The billiard player, if asked how he decides where to hit the ball, may say that he "just figures it out" but then also rubs a rabbit's foot just to make sure; and the businessman may well say that he prices at average cost, with of course some minor deviations when the market makes it necessary. The one statement is about as helpful as the other, and neither is a relevant test of the associated hypothesis.

Confidence in the maximization-of-returns hypothesis is justified by evidence of a very different character. This evidence is in part similar to that adduced on behalf of the billiard-player hypothesis – unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long. Let the apparent immediate determinant of business behavior be anything at all – habitual reaction, random chance, or whatnot. Whenever this determinant happens to lead to behavior consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not, the business will tend to lose resources and can be kept in existence only by the addition of resources from outside. The process of "natural selection" thus helps to validate the hypothesis – or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.

An even more important body of evidence for the maximization-of-returns hypothesis is experience from countless applications of the hypothesis to specific problems and the repeated failure of its implications to be contradicted. This evidence is extremely hard to document; it is scattered in numerous memorandums, articles, and monographs concerned primarily with specific concrete problems rather than with submitting the hypothesis to test. Yet the continued use and acceptance of the hypothesis over a long period, and the failure of any coherent, self-consistent alternative to be developed and be widely accepted, is strong indirect testimony to its worth. The evidence *for* a hypothesis always consists of its repeated failure to be contradicted, continues to accumulate so long as the hypothesis is used, and by its very nature is difficult to document at all comprehensively. It tends to become part of the tradition and folklore of a science revealed in the tenacity with which hypotheses are held rather than in any textbook list of instances in which the hypothesis has failed to be contradicted.

IV. The Significance and Role of the "Assumptions" of a Theory

Up to this point our conclusions about the significance of the "assumptions" of a theory have been almost entirely negative: we have seen that a theory cannot be tested by the "realism" of its "assumptions" and that the very concept of the "assumptions" of a theory is surrounded with ambiguity. But, if this were all there is to it, it would be hard to explain the extensive use of the concept and the strong tendency that we all have to speak of the assumptions of a theory and to compare the assumptions of alternative theories. There is too much smoke for there to be no fire.

In methodology, as in positive science, negative statements can generally be made with greater confidence than positive statements, so I have less confidence in the following remarks on the significance and role of "assumptions" than in the preceding remarks. So far as I can see, the "assumptions of a theory" play three different, though related, positive roles: (*a*) they are often an economical mode of describing or presenting a theory; (*b*) they sometimes facilitate an indirect test of the hypothesis by its implications; and (*c*), as already noted, they are sometimes a convenient means of specifying the conditions under which the theory is expected to be valid. The first two require more extensive discussion.

A. The Use of "Assumptions" in Stating a Theory

The example of the leaves illustrates the first role of assumptions. Instead of saying that leaves seek to maximize the sunlight they receive, we could state the equivalent hypothesis, without any apparent assumptions, in the form of a list of rules for predicting the density of leaves: If a tree stands in a level field with no other trees or other bodies obstructing the rays of the sun, then the density of leaves will tend to be such and such; if a tree is on the northern slope of a hill in the midst of a forest of similar trees, then . . . ; etc. This is clearly a far less economical presentation of the hypothesis than the statement that leaves seek to maximize the sunlight each receives. The latter statement is, in effect, a simple summary of the rules in the above list, even if the list were indefinitely extended, since it indicates both how to determine the features of the environment that are important for the particular problem and how to evaluate their effects. It is more compact and at the same time no less comprehensive.

More generally, a hypothesis or theory consists of an assertion that certain forces are, and by implication others are not, important for a particular class of phenomena and a specification of the manner of action of the forces it asserts to be important. We can regard the hypothesis as consisting of two parts: first, a conceptual world or abstract model simpler than the "real world" and containing only the forces that the hypothesis asserts to be important; second, a set of rules defining the class of phenomena for which the "model" can be taken to be an adequate representation of the "real world" and specifying the correspondence between the variables or entities in the model and observable phenomena.

These two parts are very different in character. The model is abstract and complete; it is an "algebra" or "logic." Mathematics and formal logic come into their own in checking its consistency and completeness and exploring its implications. There is no place in the model for, and no function to be served by, vagueness, maybe's, or approximations. The air pressure is zero, not "small," for a vacuum; the demand curve for the product of a competitive producer is horizontal (has a slope of zero), not "almost horizontal."

The rules for using the model, on the other hand, cannot possibly be abstract and complete. They must be concrete and in consequence incomplete – completeness is possible only in a conceptual world, not in the "real world," however that may be interpreted. The model is the logical embodiment of the half-truth, "There is nothing new under the sun"; the rules for applying it cannot neglect the equally significant half-truth, "History never repeats itself." To a considerable extent the rules can be formulated explicitly – most easily, though even then not completely, when the theory is part of an explicit more general theory as in the example of the vacuum theory for falling bodies. In seeking to make a science as "objective" as possible, our aim should be to formulate the rules explicitly in so far as possible and continually to widen the range of phenomena for which it is possible to do so. But, no matter how successful we may be in this attempt, there inevitably will remain room for judgment in applying the rules. Each occurrence has some features peculiarly its own, not covered by the explicit rules. The capacity to judge that these are or are not to be disregarded, that they should or should not affect what observable phenomena are to be identified with what entities in the model, is something that cannot be taught; it can be learned but only by experience and exposure in the "right" scientific atmosphere, not by rote. It is at this point that the "amateur" is separated from the "professional" in all sciences and that the thin line is drawn which distinguishes the "crackpot" from the scientist.

A simple example may perhaps clarify this point. Euclidean geometry is an abstract model, logically complete and consistent. Its entities are precisely defined – a line is not a geometrical figure "much" longer than it is wide or deep; it is a figure whose width and depth are zero. It is also obviously "unrealistic." There are no such things in "reality" as Euclidean points or lines or surfaces. Let us apply this abstract model to a mark made on a blackboard by a piece of chalk. Is the mark to be identified with a Euclidean line, a Euclidean surface, or a Euclidean solid? Clearly, it can appropriately be identified with a line if it is being used to represent, say, a demand curve. But it cannot be so identified if it is being used to color, say, countries on a map, for that would imply that the map would never be colored; for this purpose, the same mark must be identified with a surface. But it cannot be so identified by a manufacturer of chalk, for that would imply that no chalk would ever be used up; for his purposes, the same mark must be identified with a volume. In this simple example these judgments will command general agreement. Yet it seems obvious that, while general considerations can be formulated to guide such judgments, they can never be comprehensive and cover every possible instance; they cannot have the self-contained coherent character of Euclidean geometry itself.

In speaking of the "crucial assumptions" of a theory, we are, I believe, trying to state the key elements of the abstract model. There are generally many different ways of describing the model completely – many different sets of "postulates" which both imply and are implied by the model as a whole. These are all logically equivalent: what are regarded as axioms or postulates of a model from one point of view can be regarded as theorems from another, and conversely. The particular "assumptions" termed "crucial" are selected on grounds of their convenience in some such respects as simplicity or economy in describing the model, intuitive plausibility, or capacity to suggest, if only by implication, some of the considerations that are relevant in judging or applying the model.

B. The Use of "Assumptions" as an Indirect Test of Theory

In presenting any hypothesis, it generally seems obvious which of the series of statements used to expound it refer to assumptions and which to implications; yet this distinction is not easy to define rigorously. It is not, I believe, a characteristic of the hypothesis as such but rather of the use to which the hypothesis is to be put. If this is so, the ease of classifying statements must reflect unambiguousness in the purpose the hypothesis is designed to serve. The possibility of interchanging theorems and axioms in an abstract model implies the possibility of interchanging "implications" and "assumptions" in the substantive hypothesis corresponding to the abstract model, which is not to say that any implication can be interchanged with any assumption but only that there may be more than one set of statements that imply the rest.

For example, consider a particular proposition in the theory of oligopolistic behavior. If we assume (*a*) that entrepreneurs seek to maximize their returns by any means including acquiring or extending monopoly power, this will imply (*b*) that, when demand for a "product" is geographically unstable, transportation costs are significant, explicit price agreements illegal, and the number of producers of the product relatively small, they will tend to establish basing-point pricing systems.[17] The assertion (*a*) is regarded as an assumption and (*b*) as an implication because we accept the prediction of market behavior as the purpose of the analysis. We shall regard the assumption as acceptable if we find that the conditions specified in (*b*) are generally associated with basing-point pricing, and conversely. Let us now change our purpose to deciding what cases to prosecute under the Sherman Antitrust Law's prohibition of a "conspiracy in restraint of trade." If we now assume (*c*) that basing-point pricing is a deliberate construction to facilitate collusion under the conditions specified in *(b)*, this will imply (*d*) that entrepreneurs who participate in basing-point pricing are engaged in a "conspiracy in restraint of trade." What was formerly an assumption now becomes an implication, and conversely. We shall now regard the assumption (*c*) as valid if we find that, when entrepreneurs participate in basing-point pricing, there generally tends to be other evidence, in the form of letters, memorandums, or the like, of what courts regard as a "conspiracy in restraint of trade."

Suppose the hypothesis works for the first purpose, namely, the prediction of market behavior. It clearly does not follow that it will work for the second purpose, namely, predicting whether there is enough evidence of a "conspiracy in restraint of trade" to justify court action. And, conversely, if it works for the second purpose, it does not follow that it will work for the first. Yet, in the absence of other evidence, the success of the hypothesis for one purpose – in explaining one class of phenomena – will give us greater confidence than we would otherwise have that it may succeed for another purpose – in explaining another class of phenomena. It is much harder to say how much greater confidence it justifies. For this depends on how closely related we judge the two classes of phenomena to be, which itself depends in a complex way on similar kinds of indirect evidence, that is, on our experience in other connections in explaining by single theories phenomena that are in some sense similarly diverse.

To state the point more generally, what are called the assumptions of a hypothesis can be used to get some indirect evidence on the acceptability of the hypothesis in so far as the assumptions can themselves be regarded as implications of the hypothesis, and hence their conformity with reality as a failure of some implications to be contradicted, or in so far as the assumptions may call to mind other implications of the hypothesis susceptible to casual empirical observation.[18] The reason this evidence is indirect is that the assumptions or associated implications generally refer to a class of phenomena different from the class which the hypothesis is designed to explain; indeed, as is implied above, this seems to be the chief criterion we use in deciding which statements to term "assumptions" and which to term "implications." The weight attached to this indirect evidence depends on how closely related we judge the two classes of phenomena to be.

Another way in which the "assumptions" of a hypothesis can facilitate its indirect testing is by bringing out its kinship with other hypotheses and thereby making the evidence on their validity relevant to the validity of the hypothesis in question. For example, a hypothesis is formulated for a particular class of behavior. This hypothesis can, as usual, be stated without specifying any "assumptions." But suppose it can be shown that it is equivalent to a set of assumptions including the assumption that man seeks his own interest. The hypothesis then gains indirect plausibility from the success for other classes of phenomena of hypotheses that can also be said to make this assumption; at least, what is being done here is not completely unprecedented or unsuccessful in all other uses. In effect, the statement of assumptions so as to bring out a relationship between superficially different hypotheses is a step in the direction of a more general hypothesis.

This kind of indirect evidence from related hypotheses explains in large measure the difference in the confidence attached to a particular hypothesis by people with different backgrounds. Consider, for example, the hypothesis that the extent of racial or religious discrimination in employment in a particular area or industry is closely related to the degree of monopoly in the industry or area in question; that, if the industry is competitive, discrimination will be significant only if the race or religion of employees affects either the willingness of other employees to work with them or the acceptability of the product to customers and will be uncorrelated with the prejudices of employers.[19] This hypothesis is far more likely to appeal to an economist than to a sociologist. It can be said to "assume" single-minded pursuit of pecuniary self-interest by employers in competitive industries; and this "assumption" works well in a wide variety of hypotheses in economics bearing on many of the mass phenomena with which economics deals. It is therefore likely to seem reasonable to the economist that it may work in this case as well. On the other hand, the hypotheses to which the sociologist is accustomed have a very different kind of model or ideal world, in which singleminded pursuit of pecuniary self-interest plays a much less important role. The indirect evidence available to the sociologist on this hypothesis is much less favorable to it than the indirect evidence available to the economist; he is therefore likely to view it with greater suspicion.

Of course, neither the evidence of the economist nor that of the sociologist is conclusive. The decisive test is whether the hypothesis works for the phenomena it purports to explain. But a judgment may be required before any satisfactory test of this kind has been made, and, perhaps, when it cannot be made in the near future, in which case, the judgment will have to based on the inadequate evidence available. In addition, even when such a test can be made, the background of the scientists is not irrelevant to the judgments they reach. There is never certainty in science, and the weight of evidence for or against a hypothesis can never be assessed completely "objectively." The economist will be more tolerant than the sociologist in judging conformity of the implications of the hypothesis with experience, and he will be persuaded to accept the hypothesis tentatively by fewer instances of "conformity."

V. Some Implications for Economic Issues

The abstract methodological issues we have been discussing have a direct bearing on the perennial criticism of "orthodox" economic theory as "unrealistic" as well as on the attempts that have been made to reformulate theory to meet this charge. Economics is a "dismal" science because it assumes man to be selfish and money-grubbing, "a lightning calculator of pleasures and pains, who oscillates like a homogeneous globule of desire of happiness under the impulse of stimuli that shift him about the area, but leave him intact";[20] it rests on outmoded psychology and must be reconstructed in line with each new development in psychology; it assumes men, or at least businessmen, to be "in a continuous state of 'alert,' ready to change prices and/or pricing rules whenever their sensitive intuitions . . . detect a change in demand and supply conditions";[21] it assumes markets to be perfect, competition to be pure, and commodities, labor, and capital to be homogeneous.

As we have seen, criticism of this type is largely beside the point unless supplemented by evidence that a hypothesis differing in one or another of these respects from the theory being criticized yields better predictions for as wide a range of phenomena. Yet most such criticism is not so supplemented; it is based almost entirely on supposedly directly perceived discrepancies between the "assumptions" and the "real world." A particularly clear example is furnished by the recent criticisms of the maximization-of-returns hypothesis on the grounds that businessmen do not and indeed cannot behave as the theory "assumes" they do. The evidence cited to support this assertion is generally taken either from the answers given by businessmen to questions about the factors affecting their decisions – a procedure for testing economic theories that is about on a par with testing theories of longevity by asking octogenarians how they account for their long life – or from descriptive studies of the decision-making activities of individual firms.[22] Little if any evidence is ever cited on the conformity of businessmen's actual market behavior – what they do rather than what they say they do – with the implications of the hypothesis being criticized, on the one hand, and an alternative hypothesis, on the other.

A theory or its "assumptions" cannot possibly be thoroughly "realistic" in the immediate descriptive sense so often assigned to this term. A completely "realistic" theory of the wheat market would have to include not only the conditions directly underlying the supply and demand for wheat but also the kind of coins or credit instruments used to make exchanges; the personal characteristics of wheat-traders such as the color of each trader's hair and eyes, his antecedents and education, the number of members of his family, their characteristics, antecedents, and education, etc.; the kind of soil on which the wheat was grown, its physical and chemical characteristics, the weather prevailing during the growing season; the personal characteristics of the farmers growing the wheat and of the consumers who will ultimately use it; and so on indefinitely. Any attempt to move very far in achieving this kind of "realism" is certain to render a theory utterly useless.

Of course, the notion of a completely realistic theory is in part a straw man. No critic of a theory would accept this logical extreme as his objective; he would say that the "assumptions" of the theory being criticized were "too" unrealistic and that his objective was a set of assumptions that were "more" realistic though still not completely and slavishly so. But so long as the test of "realism" is the directly perceived descriptive accuracy of the "assumptions" – for example, the observation that "businessmen do not appear to be either as avaricious or as dynamic or as logical as marginal theory portrays them"[23] or that "it would be utterly impractical under present conditions for the manager of a multiprocess plant to attempt . . . to work out and equate marginal costs and marginal revenues for each productive factor"[24] – there is no basis for making such a distinction, that is, for stopping short of the straw man depicted in the preceding paragraph. What is the criterion by which to judge whether a particular departure from realism is or is not acceptable? Why is it more "unrealistic" in analyzing business behavior to neglect the magnitude of businessmen's costs than the color of their eyes? The obvious answer is because the first makes more difference to business behavior than the second; but there is no way of knowing that this is so simply by observing that businessmen do have costs of different magnitudes and eyes of different color. Clearly it can only be known by comparing the effect on the discrepancy between actual and predicted behavior of taking the one factor or the other into account. Even the most extreme proponents of realistic assumptions are thus necessarily driven to reject their own criterion and to accept the test by prediction when they classify alternative assumptions as more or less realistic.[25]

The basic confusion between descriptive accuracy and analytical relevance that underlies most criticisms of economic theory on the grounds that its assumptions are unrealistic as well as the plausibility of the views that lead to this confusion are both strikingly illustrated by a seemingly innocuous remark in an article on business-cycle theory that "economic phenomena are varied and complex, so any comprehensive theory of the business cycle that can apply closely to reality must be very complicated."[26] A fundamental hypothesis of science is that appearances are deceptive and that there is a way of looking at or interpreting or organizing the evidence that will reveal superficially disconnected and diverse phenomena to be manifestations of a more fundamental and relatively simple structure. And the test of this hypothesis, as of any other, is its fruits – a test that science has so far met with dramatic success. If a class of "economic phenomena" appears varied and complex, it is, we must suppose, because we have no adequate theory to explain them. Known facts cannot be set on one side; a theory to apply "closely to reality," on the other. A theory is the way we perceive "facts," and we cannot perceive "facts" without a theory. Any assertion that economic phenomena *are* varied and complex denies the tentative state of knowledge that alone makes scientific activity meaningful; it is in a class with John Stuart Mill's justly ridiculed statement that "happily, there is nothing in the laws of value which remains [1848] for the present or any future writer to clear up; the theory of the subject is complete."[27]

The confusion between descriptive accuracy and analytical relevance has led not only to criticisms of economic theory on largely irrelevant grounds but also to misunderstanding of economic theory and misdirection of efforts to repair supposed defects. "Ideal types" in the abstract model developed by economic theorists have been regarded as strictly descriptive categories intended to correspond directly and fully to entities in the real world independently of the purpose for which the model is being used. The obvious discrepancies have led to necessarily unsuccessful attempts to construct theories on the basis of categories intended to be fully descriptive.

This tendency is perhaps most clearly illustrated by the interpretation given to the concepts of "perfect competition" and "monopoly" and the development of the theory of "monopolistic" or "imperfect competition." Marshall, it is said, assumed "perfect competition"; perhaps there once was such a thing. But clearly there is no longer, and we must therefore discard his theories. The reader will search long and hard – and I predict unsuccessfully – to find in Marshall any explicit assumption about perfect competition or any assertion that in a descriptive sense the world is composed of atomistic firms engaged in perfect competition. Rather, he will find Marshall saying: "At one extreme are world markets in which competition acts directly from all parts of the globe; and at the other those secluded markets in which all direct competition from afar is shut out, though indirect and transmitted competition may make itself felt even in these; and about midway between these extremes lie the great majority of the markets which the economist and the business man have to study."[28] Marshall took the world as it is; he sought to construct an "engine" to analyze it, not a photographic reproduction of it.

In analyzing the world as it is, Marshall constructed the hypothesis that, for many problems, firms could be grouped into "industries" such that the similarities among the firms in each group were more important than the differences among them. These are problems in which the important element is that a group of firms is affected alike by some stimulus – a common change in the demand for their products, say, or in the supply of factors. But this will not do for all problems: the important element for these may be the differential effect on particular firms.

The abstract model corresponding to this hypothesis contains two "ideal" types of firms: atomistically competitive firms, grouped into industries, and monopolistic firms. A firm is competitive if the demand curve for its output is infinitely elastic with respect to its own price for some price and all outputs, given the prices charged by all other firms; it belongs to an "industry" defined as a group of firms producing a single "product." A "product" is defined as a collection of units that are perfect substitutes to purchasers so the elasticity of demand for the output of one firm with respect to the price of another firm in the same industry is infinite for some price and some outputs. A firm is monopolistic if the demand curve for its output is not infinitely elastic at some price for all outputs.[29] If it is a monopolist, the firm is the industry.[30]

As always, the hypothesis as a whole consists not only of this abstract model and its ideal types but also of a set of rules, mostly implicit and suggested by example, for identifying actual firms with one or the other ideal type and for classifying firms into industries. The ideal types are not intended to be descriptive; they are designed to isolate the features that are crucial for a particular problem. Even if we could estimate directly and accurately the demand curve for a firm's product, we could not proceed immediately to classify the firm as perfectly competitive or monopolistic according as the elasticity of the demand curve is or is not infinite. No observed demand curve will ever be precisely horizontal, so the estimated elasticity will always be finite. The relevant question always is whether the elasticity is "sufficiently" large to be regarded as infinite, but this is a question that cannot be answered, once for all, simply in terms of the numerical value of the elasticity itself, any more than we can say, once for all, whether an air pressure of 15 pounds per square inch is "sufficiently" close to zero to use the formula $s = 1/2\ gt^2$. Similarly, we cannot compute cross-elasticities of demand and then classify firms into industries according as there is a "substantial gap in the cross-elasticities of demand." As Marshall says, "The question where the lines of division between different commodities [i.e., industries] should be drawn must be settled by convenience of the particular discussion."[31] Everything depends on the problem; there is no inconsistency in regarding the same firm as if it were a perfect competitor for one problem, and a monopolist for another, just as there is none in regarding the same chalk mark as a Euclidean line for one problem, a Euclidean surface for a second, and a Euclidean solid for a third. The size of the elasticity and cross-elasticity of demand, the number of firms producing physically similar products, etc., are all relevant because they are or may be among the variables used to define the correspondence between the ideal and real entities in a particular problem and to specify the circumstances under which the theory holds sufficiently well; but they do not provide, once for all, a classification of firms as competitive or monopolistic.

An example may help to clarify this point. Suppose the problem is to determine the effect on retail prices of cigarettes of an increase, expected to be permanent, in the federal cigarette tax. I venture to predict that broadly correct results will be obtained by treating cigarette firms as if they were producing an identical product and were in perfect competition. Of course, in such a case, "some convention must be made as to the" number of Chesterfield cigarettes "which are taken as equivalent" to a Marlboro.[32]

On the other hand, the hypothesis that cigarette firms would behave as if they were perfectly competitive would have been a false guide to their reactions to price control in World War II, and this would doubtless have been recognized before the event. Costs of the cigarette firms must have risen during the war. Under such circumstances perfect competitors would have reduced the quantity offered for sale at the previously existing price. But, at that price, the wartime rise in the income of the public presumably increased the quantity demanded. Under conditions of perfect competition strict adherence to the legal price would therefore imply not only a "shortage" in the sense that quantity demanded exceeded quantity supplied but also an absolute decline in the number of cigarettes produced. The facts contradict this particular implication: there was reasonably good adherence to maximum cigarette prices, yet the quantities produced increased substantially. The common force of increased costs presumably operated less strongly than the disruptive force of the desire by each firm to keep its share of the market, to maintain the value and prestige of its brand name, especially when the excess-profits tax shifted a large share of the costs of this kind of advertising to the government. For this problem the cigarette firms cannot be treated *as if* they were perfect competitors.

Wheat farming is frequently taken to exemplify perfect competition. Yet, while for some problems it is appropriate to treat cigarette producers as if they comprised a perfectly competitive industry, for some it is not appropriate to treat wheat producers as if they did. For example, it may not be if the problem is the differential in prices paid by local elevator operators for wheat.

Marshall's apparatus turned out to be most useful for problems in which a group of firms is affected by common stimuli, and in which the firms can be treated *as if* they were perfect competitors. This is the source of the misconception that Marshall "assumed" perfect competition in some descriptive sense. It would be highly desirable to have a more general theory than Marshall's, one that would cover at the same time both those cases in which differentiation of product or fewness of numbers makes an essential difference and those in which it does not. Such a theory would enable us to handle problems we now cannot and, in addition, facilitate determination of the range of circumstances under which the simpler theory can be regarded as a good enough approximation. To perform this function, the more general theory must have content and substance; it must have implications susceptible to empirical contradiction and of substantive interest and importance.

The theory of imperfect or monopolistic competition developed by Chamberlin and Robinson is an attempt to construct such a more general theory.[33] Unfortunately, it possesses none of the attributes that would make it a truly useful general theory. Its contribution has been limited largely to improving the exposition of the economics of the individual firm and thereby the derivation of implications of the Marshallian model, refining Marshall's monopoly analysis, and enriching the vocabulary available for describing industrial experience.

The deficiencies of the theory are revealed most clearly in its treatment of, or inability to treat, problems involving groups of firms – Marshallian "industries." So long as it is insisted that differentiation of product is essential – and it is the distinguishing feature of the theory that it does insist on this point – the definition of an industry in terms of firms producing an identical product cannot be used. By that definition each firm is a separate industry. Definition in terms of "close" substitutes or a "substantial" gap in cross-elasticities evades the issue, introduces fuzziness and undefinable terms into the abstract model where they have no place, and serves only to make the theory analytically meaningless – "close" and "substantial" are in the same category as a "small" air pressure.[34] In one connection Chamberlin implicitly defines an industry as a group of firms having identical cost and demand curves.[35] But this, too, is logically meaningless so long as differentiation of product is, as claimed, essential and not to be put aside. What does it mean to say that the cost and demand curves of a firm producing bulldozers are identical with those of a firm producing hairpins?[36] And if it is meaningless for bulldozers and hairpins, it is meaningless also for two brands of toothpaste – so long as it is insisted that the difference between the two brands is fundamentally important.

The theory of monopolistic competition offers no tools for the analysis of an industry and so no stopping place between the firm at one extreme and general equilibrium at the other.[37] It is therefore incompetent to contribute to the analysis of a host of important problems: the one extreme is too narrow to be of great interest; the other, too broad to permit meaningful generalizations.[38]

VI. Conclusion

Economics as a positive science is a body of tentatively accepted generalizations about economic phenomena that can be used to predict the consequences of changes in circumstances. Progress in expanding this body of generalizations, strengthening our confidence in their validity, and improving the accuracy of the predictions they yield is hindered not only by the limitations of human ability that impede all search for knowledge but also by obstacles that are especially important for the social sciences in general and economics in particular, though by no means peculiar to them. Familiarity with the subject matter of economics breeds contempt for special knowledge about it. The importance of its subject matter to everyday life and to major issues of public policy impedes objectivity and promotes confusion between scientific analysis and normative judgment. The necessity of relying on uncontrolled experience rather than on controlled experiment makes it difficult to produce dramatic and clear-cut evidence to justify the acceptance of tentative hypotheses. Reliance on uncontrolled experience does not affect the fundmental methodological principle that a hypothesis can be tested only by the conformity of its implications or predictions with observable phenomena; but it does render the task of testing hypotheses more difficult and gives greater scope for confusion about the methodological principles involved. More than other scientists, social scientists need to be self-conscious about their methodology.

One confusion that has been particularly rife and has done much damage is confusion about the role of "assumptions" in economic analysis. A meaningful scientific hypothesis or theory typically asserts that certain forces are, and other forces are not, important in understanding a particular class of phenomena. It is frequently convenient to present such a hypothesis by stating that the phenomena it is desired to predict behave in the world of observation *as if* they occurred in a hypothetical and highly simplified world containing only the forces that the hypothesis asserts to be important. In general, there is more than one way to formulate such a description – more than one set of "assumptions" in terms of which the theory can be presented. The choice among such alternative assumptions is made on the grounds of the resulting economy, clarity, and precision in presenting the hypothesis; their capacity to bring indirect evidence to bear on the validity of the hypothesis by suggesting some of its implications that can be readily checked with observation or by bringing out its connection with other hypotheses dealing with related phenomena; and similar considerations.

Such a theory cannot be tested by comparing its "assumptions" directly with "reality." Indeed, there is no meaningful way in which this can be done. Complete "realism" is clearly unattainable, and the question whether a theory is realistic "enough" can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories. Yet the belief that a theory can be tested by the realism of its assumptions independently of the accuracy of its predictions is widespread and the source of much of the perennial criticism of economic theory as unrealistic. Such criticism is largely irrelevant, and, in consequence, most attempts to reform economic theory that it has stimulated have been unsuccessful.

The irrelevance of so much criticism of economic theory does not of course imply that existing economic theory deserves any high degree of confidence. These criticisms may miss the target, yet there may be a target for criticism. In a trivial sense, of course, there obviously is. Any theory is necessarily provisional and subject to change with the advance of knowledge. To go beyond this platitude, it is necessary to be more specific about the content of "existing economic theory" and to distinguish among its different branches; some parts of economic theory clearly deserve more confidence than others. A comprehensive evaluation of the present state of positive economics, summary of the evidence bearing on its validity, and assessment of the relative confidence that each part deserves is clearly a task for a treatise or a set of treatises, if it be possible at all, not for a brief paper on methodology.

About all that is possible here is the cursory expression of a personal view. Existing relative price theory, which is designed to explain the allocation of resources among alternative ends and the division of the product among the co-operating resources and which reached almost its present form in Marshall's *Principles of Economics*, seems to me both extremely fruitful and deserving of much confidence for the kind of economic system that characterizes Western nations. Despite the appearance of considerable controversy, this is true equally of existing static monetary theory, which is designed to explain the structural or secular level of absolute prices, aggregate output, and other variables for the economy as a whole and which has had a form of the quantity theory of money as its basic core in all of its major variants from David Hume to the Cambridge School to Irving Fisher to John Maynard Keynes. The weakest and least satisfactory part of current economic theory seems to me to be in the field of monetary dynamics, which is concerned with the process of adaptation of the economy as a whole to changes in conditions and so with short-period fluctuations in aggregate activity. In this field we do not even have a theory that can appropriately be called "the" existing theory of monetary dynamics.

Of course, even in relative price and static monetary theory there is enormous room for extending the scope and improving the accuracy of existing theory. In particular, undue emphasis on the descriptive realism of "assumptions" has contributed to neglect of the critical problem of determining the limits of validity of the various hypotheses that together constitute the existing economic theory in these areas. The abstract models corresponding to these hypotheses have been elaborated in considerable detail and greatly improved in rigor and precision. Descriptive material on the characteristics of our economic system and its operations have been amassed on an unprecedented scale. This is all to the good. But, if we are to use effectively these abstract models and this descriptive material, we must have a comparable exploration of the criteria for determining what abstract model it is best to use for particular kinds of problems, what entities in the abstract model are to be identified with what observable entities, and what features of the problem or of the circumstances have the greatest effect on the accuracy of the predictions yielded by a particular model or theory.

Progress in positive economics will require not only the testing and elaboration of existing hypotheses but also the construction of new hypotheses. On this problem there is little to say on a formal level. The construction of hypotheses is a creative act of inspiration, intuition, invention; its essence is the vision of something new in familiar material. The process must be discussed in psychological, not logical, categories; studies in autobiographies and biographies, not treatises on scientific method; and promoted by maxim and example, not syllogism or theorem.

Notes

1. (London: Macmillan & Co., 1981), pp. 34–5 and 46.
2. Social science or economics is by no means peculiar in this respect – witness the importance of personal beliefs and of "home" remedies in medicine wherever obviously convincing evidence for "expert" opinion is lacking. The current prestige and acceptance of the views of physical scientists in their fields of specialization – and, all too often, in other fields as well – derives, not from faith alone, but from the evidence of their works, the success of their predictions, and the dramatic achievements from applying their results. When economics seemed to provide such evidence of its worth, in Great Britain in the first half of the nineteenth century, the prestige and acceptance of "scientific economics" rivaled the current prestige of the physical sciences.
3. The interaction between the observer and the process observed that is so prominent a feature of the social sciences, besides its more obvious parallel in the

physical sciences, has a more subtle counterpart in the indeterminacy principle arising out of the interaction between the process of measurement and the phenomena being measured. And both have a counterpart in pure logic in Gödel's theorem, asserting the impossibility of a comprehensive self-contained logic. It is an open question whether all three can be regarded as different formulations of an even more general principle.

4. One rather more complex example is stabilization policy. Superficially, divergent views on this question seem to reflect differences in objectives; but I believe that this impression is misleading and that at bottom the different views reflect primarily different judgments about the source of fluctuations in economic activity and the effect of alternative countercyclical action. For one major positive consideration that accounts for much of the divergence see "The Effects of a Full-Employment Policy on Economic Stability: A Formal Analysis," *infra*, pp. 117–32. For a summary of the present state of professional views on this question see "The Problem of Economic Instability," a report of a subcommittee of the Committee on Public Issues of the American Economic Association, *American Economic Review*, XL (September, 1950), 501–38.
5. Final quoted phrase from Alfred Marshall, "The Present Position of Economics" (1885), reprinted in *Memorials of Alfred Marshall*, ed. A. C. Pigou (London: Macmillan & Co., 1925), p. 164. See also "The Marshallian Demand Curve," *infra*, pp. 56–7, 90–1.
6. See "Lange on Price Flexibility and Employment: A Methodological Criticism," *infra*, pp. 282–9.
7. "The Marshallian Demand Curve," *infra*, p. 57.
8. The qualification is necessary because the "evidence" may be internally contradictory, so there may be no hypothesis consistent with it. See also "Lange on Price Flexibility and Employment," *infra*, pp. 282–3.
9. See "Lange on Price Flexibility and Employment," *infra, passim.*
10. See also Milton Friedman and L. J. Savage, "The Expected-Utility Hypothesis and the Measurability of Utility," *Journal of Political Economy*, LX (December, 1952), 463–74, esp. pp. 465–7.
11. In recent years some economists, particularly a group connected with the Cowles Commission for Research in Economics at the University of Chicago, have placed great emphasis on a division of this step of selecting a hypothesis consistent with known evidence into two substeps: first, the selection of a class of admissible hypotheses from all possible hypotheses (the choice of a "model" in their terminology); second, the selection of one hypothesis from this class (the choice of a "structure"). This subdivision may be heuristically valuable in some kinds of work, particularly in promoting a systematic use of available statistical evidence and theory. From a methodological point of view, however, it is an entirely arbitrary subdivision of the process of deciding on a particular hypothesis that is on a par with many other subdivisions that may be convenient for one purpose or another or that may suit the psychological needs of particular investigators.

    One consequence of this particular subdivision has been to give rise to the so-called "identification" problem. As noted above, if one hypothesis is consistent with available evidence, an infinite number are. But while this is true for the class of hypotheses as a whole, it may not be true of the subclass obtained in the first of the above two steps – the "model." It may be that the evidence to be used to select the final hypothesis from the subclass can be consistent with at most one hypothesis in it, in which case the "model" is said to be "identified"; otherwise it is said to be "unidentified." As is clear from this way of describing the concept of "identification," it is essentially a special case of the more general problem of selecting among the alternative hypotheses equally consistent with the evidence – a problem that must be decided by some such arbitrary principle as Occam's razor. The introduction of two substeps in selecting a hypothesis makes this problem arise at the two corresponding stages and gives it a special cast. While the class of all hypotheses is always unidentified, the subclass in a "model" need not be, so the problem arises of conditions that a "model" must satisfy to be identified. However useful the two substeps may be in some contexts, their introduction raises the danger that different criteria will unwittingly be used in making the same kind of choice among alternative hypotheses at two different stages.

On the general methodological approach discussed in this footnote see Tryvge Haavelmo, "The Probability Approach in Econometrics," *Econometrica*, Vol. XII (1944), Supplement; Jacob Marschak, "Economic Structure, Path, Policy, and Prediction," *American Economic Review*, XXXVII, (May, 1947), 81–84, and "Statistical Inference in Economics: An Introduction," in T. C. Koopmans (ed.), *Statistic Inference in Dynamic Economic Models* (New York: John Wiley & Sons, 1950); T. C. Koopmans, "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, XL (December, 1945), 448–66; Gershon Cooper, "The Role of Economic Theory in Econometric Models," *Journal of Farm Economics*, XXX (February, 1948), 101–16. On the identification problem see Koopmans, "Identification Problems in Econometric Model Construction," *Econometrica, XVII* (April, 1949), 125–44; Leonid Hurwicz, "Generalization of the Concept of Identification," in Koopmans (ed.), *Statistical Inference in Dynamic Economic Models.*

12. The converse of the proposition does not of course hold: assumptions that are unrealistic (in this sense) do not guarantee a significant theory.
13. See R. A. Lester, "Shortcomings of Marginal Analysis for Wage-Employment Problems," *American Economic Review*, XXXVI (March, 1946), 62–82; Fritz Machlup, "Marginal Analysis and Empirical Research," *American Economic Review*, XXXVI (September, 1946), 519–54; R. A. Lester, "Marginalism, Minimum Wages, and Labor Markets," *American Economic Review*, XXXVII (March, 1947), 135–48; Fritz Machlup, "Rejoinder to an Antimarginalist," *American Economic Review*, XXXVII (March, 1947), 148–54; G. J. Stigler, "Professor Lester and the Marginalists," *American Economic Review*, XXXVII (March, 1947), 154–57; H. M. Oliver, Jr., "Marginal Theory and Business Behavior," *American Economic Review*, XXXVII (June, 1947), 375–83; R. A. Gordon, "Short-Period Price Determination in Theory and Practice," *American Economic Review*, XXXVIII (June, 1948), 265–88.

It should be noted that, along with much material purportedly bearing on the validity of the "assumptions" of marginal theory, Lester does refer to evidence on the conformity of experience with the implications of the theory, citing the reactions of employment in Germany to the Papen plan and in the United States to changes in minimum-wage legislation as examples of lack of conformity. However, Stigler's brief comment is the only one of the other papers that refers to this evidence. It should be noted that Machlup's thorough and careful exposition of the logical structure and meaning of marginal analysis is called for by the misunderstandings on this score that mar Lester's paper and almost conceal the evidence he presents that is relevant to the key issue he raises. But, in Machlup's emphasis on the logical structure, he comes perilously close to presenting the theory as a pure tautology, though it is evident at a number of points that he is aware of this danger and anxious to avoid it. The papers by Oliver and Gordon are the most extreme in the exclusive concentration on the conformity of the behavior of businessmen with the "assumptions" of the theory.

14. This example, and some of the subsequent discussion, though independent in origin, is similar to and in much the same spirit as an example and the approach in an important paper by Armen A. Alchian, "Uncertainty, Evolution, and Economic Theory," *Journal of Political Economy*, LVIII (June, 1950), 211–21.
15. Milton Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, LVI (August, 1948), 298. Reprinted in American Economic Association, *Readings in Price Theory* (Chicago: Richard D. Irwin, Inc., 1952), pp. 57–96.
16. It seems better to use the term "profits" to refer to the difference between actual and "expected" results, between *ex post* and *ex ante* receipts. "Profits" are then a result of uncertainty and, as Alchian (*op. cit.*, p. 212), following Tintner, points out, cannot be deliberately maximized in advance. Given uncertainty, individuals or firms choose among alternative anticipated probability distributions of receipts or incomes. The specific content of a theory of choice among such distributions depends on the criteria by which they are supposed to be ranked. One hypothesis supposes them to be ranked by the mathematical expectation of utility corresponding to them (see Friedman and Savage, "The Expected-Utility Hypothesis and the Measurability of Utility," *op. cit.*). A special case of this hypothesis or an alternative to it ranks probability distribution by the mathematical expectation of the money receipts corresponding to them. The latter is perhaps more applicable, and more frequently applied, to firms than to individuals. The term "expected returns" is intended to be sufficiently broad to apply to any of these alternatives.

    The issues alluded to in this note are not basic to the methodological issues being discussed, and so are largely by-passed in the discussion that follows.
17. See George J. Stigler, "A Theory of Delivered Price Systems," *American Economic Review*, XXXIX (December, 1949), 1143–57.
18. See Friedman and Savage, "The Expected-Utility Hypothesis and the Measurability of Utility," *op. cit.*, pp. 466–7, for another specific example of this kind of indirect test.
19. A rigorous statement of this hypothesis would of course have to specify how "extent of racial or religious discrimination" and "degree of monopoly" are to be judged. The loose statement in the text is sufficient, however, for present purposes.
20. Thorstein Veblen, "Why Is Economics Not an Evolutionary Science?" (1898), reprinted in *The Place of Science in Modern Civilization* (New York, 1919), p. 73.

21. Oliver, *op. cit.*, p. 381.
22. See H. D. Henderson, "The Significance of the Rate of Interest," *Oxford Economic Papers*, No. 1 (October, 1938), pp. 1–13; J. E. Meade and P. W. S. Andrews, "Summary of Replies to Questions on Effects of Interest Rates," *Oxford Economic Papers*, No. 1 (October, 1938), pp. 14–31; R. F. Harrod, "Price and Cost in Entrepreneurs' Policy," *Oxford Economic Papers*, No. 2 (May, 1939), pp. 1–11; and R. J. Hall and C. J. Hitch, "Price Theory and Business Behavior," *Oxford Economic Papers*, No. 2 (May, 1939), pp. 12–45; Lester, "Shortcomings of Marginal Analysis for Wage-Employment Problems," *op. cit.*; Gordon, *op. cit.* See Fritz Machlup, "Marginal Analysis and Empirical Research," *op. cit.*, esp. Sec. II, for detailed criticisms of questionnaire methods.

    I do not mean to imply that questionnaire studies of businessmen's or others' motives or beliefs about the forces affecting their behavior are useless for all purposes in economics. They may be extremely valuable in suggesting leads to follow in accounting for divergencies between predicted and observed results; that is, in constructing new hypotheses or revising old ones. Whatever their suggestive value in this respect, they seem to me almost entirely useless as a means of *testing* the validity of economic hypotheses. See my comment on Albert G. Hart's paper, "Liquidity and Uncertainty," *American Economic Review*, XXXIX (May, 1949), 198–99.
23. Oliver, *op. cit.*, p. 382.
24. Lester, "Shortcomings of Marginal Analysis for Wage-Employment Problems," *op. cit.*, p. 75.
25. E.g., Gordon's direct examination of the "assumptions" leads him to formulate the alternative hypothesis generally favored by the critics of the maximization-of-returns hypothesises follows: "There is an irresistible tendency to price on the basis of average total costs for some 'normal' level of output. This is the yardstick, the short-cut, that businessmen and accountants use, and their aim is more to earn satisfactory profits and play safe than to maximize profits" (*op. cit.*, p. 275). Yet he essentially abandons this hypothesis, or converts it into a tautology, and in the process implicitly accepts the test by prediction when he later remarks: "Full cost and satisfactory profits may continue to be the objectives even when total costs are shaded to meet competition or exceeded to take advantage of a sellers' market" (*ibid.*, p. 284). Where here is the "irresistible tendency"? What kind of evidence could contradict this assertion?
26. Sidney S. Alexander, "Issues of Business Cycle Theory Raised by Mr. Hicks," *American Economic Review*, XLI (December, 1951), 872.
27. *Principles of Political Economy* (Ashley ed.; Longmans, Green & Co., 1929), p. 436.
28. *Principles*, p. 329; see also pp. 35, 100, 341, 347, 375, 546.
29. This ideal type can be divided into two types: the oligopolistic firm, if the demand curve for its output is infinitely elastic at some price for some but not all outputs; the monopolistic firm proper, if the demand curve is nowhere infinitely elastic (except possibly at an output of zero).
30. For the oligopolist of the preceding note an industry can be defined as a group of firms producing the same product.
31. *Principles*, p. 100.
32. Quoted parts from *ibid.*

33. E. H. Chamberlin, *The Theory of Monopolistic Competition* (6th ed.; Cambridge: Harvard University Press, 1950); Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan & Co., 1933).
34. See R. L. Bishop, "Elasticities, Cross-elasticities, and Market Relationships," *American Economic Review*, XLII (December, 1952), 779–803, for a recent attempt to construct a rigorous classification of market relationships along these lines. Despite its ingenuity and sophistication, the result seems to me thoroughly unsatisfactory. It rests basically on certain numbers being classified as "large" or "small," yet there is no discussion at all of how to decide whether a particular number is "large" or "small," as of course there cannot be on a purely abstract level.
35. *Op. cit.*, p. 82.
36. There always exists a transformation of quantities that will make either the cost curves or the demand curves identical; this transformation need not, however, be linear, in which case it will involve different-sized units of one product at different levels of output. There does not necessarily exist a transformation that will make both pairs of curves identical.
37. See Robert Triffin, *Monopolistic Competition and General Equilibrium Theory* (Cambridge: Harvard University Press, 1940), esp. pp. 188–89.
38. For a detailed critique see George J. Stigler, "Monopolistic Competition in Retrospect," in *Five Lectures on Economic Problems* (London: Macmillan & Co., 1949), pp. 12–14.

EIGHT

Testability and Approximation

Herbert Simon

Herbert Simon (1916–2001) was born in Milwaukee, Wisconsin, and received his Ph.D. in political science from the University of Chicago. He taught at the Illinois Institute of Technology and at Carnegie-Mellon University. Simon made major contributions to a number of different disciplines including political science, psychology, philosophy, and economics. He was awarded the Nobel Prize in economics in 1978. The following short essay was written for a symposium on Milton Friedman's methodology that was held at the 1962 meetings of the American Economic Association.

I find methodological inquiry interesting and instructive to the extent to which it addresses itself to concrete problems of empirical science. Thus, while I find myself in general agreement with almost everything that has been said in the previous papers and by discussants, I should like to pitch my remarks at a level less abstract than theirs.

The Relation of Premises and Conclusions in Economic Theory

Professor Nagel has pointed out that whether a particular proposition is a fundamental assumption of a theory or one of its derived conclusions is relative to the formulation of the theory. If this were the whole story, then asymmetry between assumptions and derivations in Friedman's position – what Professor Samuelson called the F-Twist, and what I like to think of as Friedman's "principle of unreality" – would be entirely arbitrary. Professor Krupp's remarks on composition laws and the relation of microscopic to macroscopic theories suggest, however, that something more is at issue.

Originally published as "Problems of Methodology – Discussion," by Herbert Simon, in the *American Economic Review: Papers and Proceedings*, vol. 53(1963): 229–31. Reprinted by permission of the American Economic Association.

Since the prefixes "micro" and "macro" have rather special meanings in economics, let me talk instead of theories of economic actors and theories of economic markets, respectively. In the present context, the relevant theory at the actor level can be approximated by the propositions: *X* – businessmen desire to maximize profits; *Y* – businessmen can and do make the calculations that identify the profit-maximizing course of action. The theory at the market level may be summed up as: *Z* – prices and quantities are observed at those levels which maximize the profits of the firms in the market. (For simplicity, let us assume that we mean the maximum of perfect competition theory.)

Defending the theory consisting of *X*, *Y*, and *Z*, Friedman asserts that it doesn't matter if *X* and *Y* are false, provided *Z* is true. Professors Nagel and Samuelson have already exposed the logical fallacy in using the validity of *Z* to support *X* and *Y*, or to support consequences of *X* and *Y* that do not follow from *Z* alone. But there are other equally serious difficulties in Friedman's position.

X and *Y* are taken as premises and *Z* as a conclusion is not just a matter of taste in formulation of the theory. The formulation fits our common, if implicit, notions of explanation. We explain the macroscopic by the microscopic (plus some composition laws) – the market by the actors. We do this partly because it satisfies our feeling that individual actors are the simple components of the complex market; hence proper explanatory elements. We do it partly because *X* and *Y*, plus the composition laws, allow us to derive other propositions at the market level – say, about shifting of taxes, or other policy matters – which we are not able to test by direct observation.

The logical fallacy in Friedman's principle of unreality has exerted so much fascination – both in this session and elsewhere – that attention has been distracted from its other errors. Most critics have accepted Friedman's assumption that proposition *Z* is the empirically tested one, while *X* and *Y* are not directly observable. This, of course, is nonsense. No one has, in fact, observed whether the actual positions of business firms are the profit-maximizing ones; nor has anyone proposed a method of testing this proposition by direct observation. I cannot imagine what such a test would be, since the tester would be as incapable as business firms are of discovering what the optimal position actually is.

If, under these circumstances, *Z* is a valid theory, it must be because it follows from empirically valid assumptions about actors together with empirically valid composition laws. Now we do have a considerable body of evidence about *X* and *Y*, and the vast weight of evidence with respect to *Y*, at least, is that it is false. The expressed purpose of Friedman's principle of unreality is to save classical theory in the face of the patent invalidity of Y. (The Alchian survival argument that "only profit-maximizers survive," does not help matters, since it, like Z, cannot be tested by direct observation – we cannot identify the profit-maximizers.)

The remedy for the difficulty is straightforward, although it may involve more empirical work at the level of the individual actors than most conventionally-trained economists find comfortable. Let us make the observations necessary to discover and test true propositions, call them X' and Y', to replace the false X and Y. Then let us construct a new market theory on these firmer foundations. This is not, of course, a novel proposal. The last two decades have seen it carried a long distance toward execution.

Ideal Types and Approximations

My final comment is related to the previous one. There has been much talk at this session of ideal types: perfect vacuums and perfect competition. I am not satisfied with the answers to Friedman's argument that he has as much right as the physicists to make unreal assumptions. Was Galileo also guilty of using the invalid principle of unreality? I think not. I think he was interested in behavior in perfect vacuums not because there aren't any in the real world, but because the real world sometimes sufficiently approximates them to make their postulation interesting.

Let me propose a methodological principle to replace the principle of unreality. I should like to call it the "principle of continuity of approximation." It asserts: if the conditions of the real world approximate sufficiently well the assumptions of an ideal type, the derivations from these assumptions will be approximately correct. Failure to incorporate this principle into his formulation seems to me a major weakness in the interesting approach of Professor Papandreou's paper. Unreality of premises is not a virtue in scientific theory; it is a necessary evil – a concession to the finite computing capacity of the scientist that is made tolerable by the principle of continuity of approximation.

Working scientists employ the principle of continuity all the time. Unfortunately, it has no place in modern statistical theory. The word "significant" has been appropriated by the statisticians to mean "unlikely to have arisen by chance." Now, in testing extreme hypotheses – ideal types – we do not primarily want to know whether there are deviations of observation from theory which are "significant" in this sense. It is far more important to know whether they are significant in the sense that the approximation of theory to reality is beyond the limits of our tolerance. Until this latter notion of significance has been properly formalized and incorporated in statistical methodology, we are not going to accord proper methodological treatment to extreme hypotheses. The discussion at this session has not provided the solution, but it has identified this problem as one of central methodological importance for economics.

NINE

Why Look Under the Hood?

Daniel M. Hausman

Daniel M. Hausman (1947–) is Professor of Philosophy at the University of Wisconsin–Madison. He received his Ph.D. from Columbia University, and he writes mainly on issues of economic methodology and the theory of causality.

Methodologists have had few kind words for Milton Friedman's "The Methodology of Positive Economics" [1953, Chapter 7 in this volume], yet its influence persists. Why? One answer is that methodologists have missed an important argument, which economists have found persuasive. Unlike Hirsch and de Marchi (1990), I am concerned here with the argument, not with "what Friedman really meant."

Friedman declares, "The ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed" (p. 7). This is the central thesis of instrumentalism. But from a standard instrumentalist perspective, in which *all* the observable consequences of a theory are significant, it is impossible to defend Friedman's central claim that the realism of assumptions is irrelevant to the assessment of a scientific theory. For the assumptions of economics are testable, and a standard instrumentalist would not dismiss apparent disconfirmations. Indeed, the distinction between assumptions and implications is superficial. The survey results reported by Richard Lester and others, which Friedman finds irrelevant and wrong-headed (pp. 15, 31f), are as much predictions of neoclassical theory as are claims about market phenomena.

I would like to thank John Dreher, Merton Finkler, Daniel Hammond, Erkki Koskela, Michael McPherson, and Herbert Simon for useful criticisms and suggestions.

Reprinted by permission of Cambridge University Press from *Essays on Philosophy and Economic Methodology* by Daniel M. Hausman. Cambridge: Cambridge University Press, 1992, pp. 70–3.

But, like Lawrence Boland (1979), I contend that Friedman is *not* a standard instrumentalist. Consider the following passages:

> Viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena which it is intended to "explain." (pp. 8–9)

> For this test [of predictions] to be relevant, the deduced facts must be about the class of phenomena the hypothesis is designed to explain; (pp. 12–13)

> The decisive test is whether the hypothesis works for the phenomena it purports to explain. (p. 30)[1]

Friedman *rejects* a standard instrumentalist concern with *all* the predictions of a theory. A good tool need not be an all-purpose tool. Friedman holds that the goal of economics is "narrow predictive success" – correct prediction only for "the class of phenomena the hypothesis is designed to explain." Lester's surveys are irrelevant because their results are not among the phenomena that the theory of the firm was designed to explain. On just these grounds, many economists dismiss any inquiry into whether the claims of the theory of consumer choice are true of individuals.

I suggest that Friedman uses this view that science aims at narrow predictive success as a premise in the following implicit argument:

(1) A good hypothesis provides valid and meaningful predictions concerning the class of phenomena it is intended to explain. (premise)
(2) The only test of whether an hypothesis is a good hypothesis is whether it provides valid and meaningful predictions concerning the class of phenomena it is intended to explain.[2] (invalidly from 1)
(3) Any other facts about an hypothesis, including whether its assumptions are realistic, are irrelevant to its scientific assessment. (trivially from 2)

If (1) the criterion of a good theory is narrow predictive success, then surely (2) the test of a good theory is narrow predictive success, and Friedman's claim that the realism of assumptions is irrelevant follows trivially. This is a tempting and persuasive argument.

But it is fallacious. (2) is not true and does not follow from (1). To see why, consider the following analogous argument:

(1′) A good used car drives safely, economically, and comfortably. (oversimplified premise)
(2′) The only test of whether a used car is a good used car is whether it drives safely, economically, and comfortably. (invalidly from 1′)

(3′) Anything one discovers by opening the hood and checking the separate components of a used car is irrelevant to its assessment. (trivially from 2′)

Presumably nobody believes 3.[3] What is wrong with the argument? It assumes that a road test is a conclusive test of a car's future performance. If this assumption were true, if it were possible (and cheap) to do a total check of the performance of a used car for the whole of its future, then there would indeed be no point in looking under the hood. For we would know everything about its performance, which is all we care about. But a road test only provides a small sample of this performance. Thus a mechanic who examines the engine can provide relevant and useful information. The mechanic's input is particularly important when one wants to use the car under new circumstances and when the car breaks down. Obviously one wants a sensible mechanic who notes not just that the components are used and imperfect, but who can judge how well the components are likely to serve their separate purposes.

Similarly, given Friedman's view of the goal of science, there would be no point in examining the assumptions of a theory if it were possible to do a "total" assessment of its performance with respect to the phenomena it was designed to explain. But one cannot make such an assessment. Indeed, the point of a theory is to guide us in circumstances where we do not already know whether the predictions are correct.[4] There is thus much that may be learned by examining the components (assumptions) of a theory and its "irrelevant" predictions. Such consideration of the "realism" of assumptions is particularly important when extending the theory to new circumstances or when revising it in the face of predictive failure.[5] Again what is relevant is not whether the assumptions are perfectly true, but whether they are adequate approximations and whether their falsehood is likely to matter for particular purposes. Saying this is not conceding Friedman's case. Wide, not narrow predictive success constitutes the grounds for judging whether a theory's assumptions are adequate approximations. The fact that a computer program works in a few instances does not render study of its algorithm and code superfluous or irrelevant.

There is a grain of truth in Friedman's defense of theories containing unrealistic assumptions. For *some* failures of assumptions may be irrelevant. Just as a malfunctioning air-conditioner is insignificant to a car's performance in Alaska, so is the falsity of the assumption of infinite divisibility unimportant in hypotheses concerning markets for basic grains. Given Friedman's narrow view of the goals of science (which I am conceding for the purposes of argument, but would otherwise contest), the realism of assumptions may thus sometimes be irrelevant. But this bit of practical wisdom does not support Friedman's strong conclusion that only narrow predictive success is relevant to the assessment of an hypothesis.

One should note three qualifications. First, we sometimes have a wealth of information concerning the track record of both theories and of used cars. I may know that my friend's old Mustang has been running without trouble for the past seven years. The more information we have about performance, the less important is separate examination of components. But it remains sensible to assess assumptions or components, particularly in circumstances of breakdown and when considering a new use. Second, intellectual tools, unlike mechanical tools, do not wear out. But if one has not yet grasped the fundamental laws governing a subject matter and does not fully know the scope of the laws and the boundary conditions on their validity, then generalizations are as likely to break down as are physical implements. Third (as Erkki Koskela reminded me), it is easier to interpret a road test than an econometric study. The difficulties of testing in economics make it all the more mandatory to look under the hood.

When either theories or used cars work, it makes sense to use them – although caution is in order if their parts have not been examined or appear to be faulty. But known performance in some sample of their given tasks is not the only information relevant to an accurate assessment of either. Economists must (and do) look under the hoods of their theoretical vehicles. When they find embarrassing things there, they must not avert their eyes and claim that what they have found cannot matter. *Even if all one cares about is predictive success in some limited domain, one should still be concerned about the realism of the assumptions of an hypothesis and the truth of its irrelevant or unimportant predictions.*

Notes

1. See also [Friedman (1953)], pp. 15, 20, and 41.
2. Notice that (2) does not say that the only test of a hypothesis is whether its predictions are valid. It says that the only test is the validity of only *some* of its predictions, namely those concerning "the class of phenomena the hypothesis is intended to explain." This is overstated, and (I repeat) I am not concerned to provide the best interpretation of Friedman's whole methodology. In his essay Friedman concedes a role for assumptions in facilitating an "indirect" test of a theory: "Yet, in the absence of other evidence, the success of the hypothesis for one purpose – in explaining one class of phenomena – will give us greater confidence

than we would otherwise have that it may succeed for another purpose – in explaining another class of phenomena. It is much harder to say how much greater confidence it justifies. For this depends on how closely related we judge the two classes of phenomena to be" (p. 28). The last sentence still limits the relevance of the correctness of predictions concerning phenomena that are remote from those that the theory is designed to explain, and Friedman clearly believes that the evidential force of indirect tests is much less than that of tests concerning the range of phenomena that the theory is intended to "explain." Daniel Hammond (unpublished) has argued that these qualifications were not part of the original draft of the essay.

3. Those who do should get in touch. I've got some fine old cars for you at bargain prices.
4. Friedman partially recognizes this point when he writes (according to Hammond, echoing criticisms George Stigler and Arthur Burns offered of an earlier draft), "The decisive test is whether the hypothesis works for the phenomena it purports to explain. But a judgment may be required before any satisfactory test of this kind has been made, and, perhaps, when it cannot be made in the near future, in which case, the judgment will have to be based on the inadequate evidence available" (1953, p. 30).
5. With what seems to me inconsistent good sense, Friedman again partly recognizes the point, "I do not mean to imply that questionnaire studies of businessmen's or other's motives or beliefs about the forces affecting their behavior are useless for all purposes in economics. They may be extremely valuable in suggesting leads to follow in accounting for divergences between predicted and observed results; that is, in constructing new hypotheses or revising old ones. Whatever their suggestive value in this respect, they seem to me almost entirely useless as a means of *testing* the validity of economic hypotheses" (1953, p. 31n).

TEN

Popper and Lakatos in Economic Methodology[1]

D. Wade Hands

D. Wade Hands (1951–) was educated at the University of Houston and then Indiana University and has taught at the University of Puget Sound since 1980. He is one of the leading figures in contemporary economic methodology. Hands was President of the History of Science Society in 2005–2006 and is currently the editor of *The Journal of Economic Methodology*. His most important book, *Reflection without Rules*, won the Spengler Book Prize from the History of Economic Society in 2004. This essay provides a brief introduction to the ideas of Karl Popper and Imre Lakatos and to the issues that arise in applying them to the philosophical understanding of economics.

Overview

The purpose of this chapter is to critically reappraise the methodological advice offered to economists by Popperian philosophy, in particular Popperian falsificationism and Lakatos's 'methodology of scientific research programmes'. These two philosophical positions and the difficulties they raise for economic methodology are carefully considered in the chapter. It is argued that while economists have benefited from the influence of Popperian philosophy in a number of ways, neither falsificationism nor Lakatos's methodology provide an appropriate guide to the acceptance or rejection of economic theories. The implications and caveats surrounding this argument are considered in the conclusion.

Pages 61–75 of *Rationality, Institutions and Economic Methodology*, edited by Uskali Mäki, Bo Gustafsson, and Christian Knudsen. London: Routledge.

Introduction

Popperian philosophy of science has been extremely influential in economic methodology. Popperian 'falsificationism', first introduced into economics by Hutchison (1938), remains one of the dominant approaches to economic methodology. In addition to this direct influence, Popperian philosophy has also affected economic methodology through the work of Imre Lakatos. A fairly extensive literature has developed around the question of the applicability of Lakatos's 'methodology of scientific research programmes' (MSRP) to economics.[2]

The purpose of this chapter is to critically reappraise the methodological advice given by Popperian philosophy. In this reappraisal both Popperian falsificationism and Lakatos's MSRP will be examined. Neo-institutionalist economics will not be explicitly discussed; instead the focus will be the general standards for economic theory choice which influence every economic theory (including neo-institutionalism). Throughout the discussion the philsophical positions will be appraised only with respect to *economic methodology*: 'economic' in that only economics and not other fields of enquiry will be discussed, and 'methodology' in that only questions of theory choice and theory appraisal (not more general philosophical considerations) will be examined. In particular, questions such as whether 'economic methodology' should be pursued at all (recently raised by McCloskey (1985)) will not be examined here.

Falsificationism

No doubt Karl Popper is best known for his falsificationist approach to the philosophy of science: a theory first presented in *Logik der Forschung* in 1934 (English translation, Popper 1959). Falsificationism represents Popper's view of the growth of scientific knowledge as well as his solution to (or dissolution of) the problem of induction. It is for falsificationism that Popper claims responsibility for the death of logical positivism (Popper 1976b: 88).

Popperian falsificationism is actually composed of two separate theses: one on demarcation (demarcating science from non-science) and one on methodology (how science should be practised). The demarcation thesis is that for a theory to be 'scientific' it must be at least potentially falsifiable by empirical observation, that is, there must exist at least one empirical basic statement which is in conflict with the theory.[3] This potential falsifiability is a logical relationship between the theory and a basic statement; in particular, the demarcation criterion only requires that it be logically possible to falsify the theory, not that such a falsification has ever been attempted.[4] While Popper's demarcation criterion has been the subject of an extensive debate in the philosophical literature, demarcation is seldom the issue in economics. For economists the more important issue is methodology (choosing between/among theories not merely labelling them scientific or unscientific) and Popperian *methodology* requires the practical (not just logical) falsifiability of scientific theories.

In a nutshell, falsificationist scientific practice proceeds as follows. The scientist starts with a scientific problem situation (something requiring a scientific explanation) and proposes a bold conjecture which might offer a solution to the problem. Next the conjecture is severely tested by comparing its least likely consequences with the relevant empirical data. Popper's argument for severe testing is that a test will be *more severe* the more prima facie *unlikely* the consequence that is being tested; the theory should be forced to 'stick its neck out', to 'offer the enemy, namely nature, the most exposed and extended surface' (Gellner 1974: 171). The final step in the falsificationist game depends on how the theory has performed during the testing stage. If the implications of the theory are not consistent with the evidence, then the conjecture is falsified and it should be replaced by a new conjecture which is *not ad hoc* relative to the original, that is, the new conjecture should not be contrived solely to avoid this empirical anomaly.[5] If the theory is not falsified by the evidence then it is considered corroborated and it is accepted provisionally. Given Popper's fallibilism this acceptance is provisional forever; the method does not necessarily result in true theories, only ones that have faced a tough empirical opponent and won.

Now while there are a number of reasons why economists have felt that Popperian falsificationism would be a desirable methodology, the fact is that *falsificationism is seldom if ever practised in economics*. This seems to be the one point generally agreed upon by recent methodological commentators. In fact, this (empirical) claim is supported at length by the case studies in Blaug (1980), a book which consistently advocates falsificationism as a normative ideal. The disagreement between critics and defenders of falsificationism is *not whether it has been practised*, basically it has not, but rather *whether it should be practised*. The real questions are whether the profession should 'try harder' to practise falsificationism though it has failed to do so in the past, and the related question of whether the discipline of economics would be substantially improved by a conscientious falsificationist practice.

One approach to the question of the appropriateness of falsificationism in economics would be to directly address the question of the adequacy of Popper's falsificationist methodology as a *general* approach to the growth of scientific knowledge; this is not the approach that will be followed here. Rather than delving into this general question, the following discussion will simply survey some of the criticisms which falsificationism has received explicitly as an economic methodology. This list of criticisms is not exhaustive, but it does capture the major concerns which have been raised regarding the falsificationism in economics. The list is not necessarily in order of importance.[6]

1. For a number of reasons, the so-called Duhemian problem (or Duhem–Quine problem) presents a great difficulty in economics.[7] First, the complexity of human behaviour requires the use of numerous initial conditions and strong simplifying assumptions. Some of these restrictions may actually be false (such as the infinite divisibility of commodities), some of these assumptions may be logically unfalsifiable (such as the assumptions of eventually diminishing returns), while still others may be logically falsifiable but practically unfalsifiable (such as the completeness assumption in consumer choice theory). Even where assumptions and restrictions can be tested, such testing is very difficult because of the absence of a suitably controlled laboratory environment.[8] In the presence of such a variety of restrictions it is virtually impossible to 'aim the arrow of *modus tollens*' at one particular problematic element of the set auxiliary hypotheses when contrary evidence is found. Second, there are many questions and disagreements about the empirical basis in economics. It is always possible to argue that what was observed was 'not really' involuntary unemployment or 'not really' economic profit, etc. Although it is fundamental to Popperian philosophy that the empirical basis need not be incorrigible, it is necessary that there be a generally accepted convention regarding the empirical basis,[9] and in economics even such conventions are often not available. Third, even if these first two problems have somehow been eliminated it is still possible for the social sciences to have feedback effects that do not exist in the physical sciences. The test of an economic theory may itself alter the initial conditions for the test. Conducting a test of the relationship between the money supply and the price level may alter expectations in such a way that the initial conditions (which were true 'initially') are not true after the test (or if the 'same' test were conducted again).

2. Related to, but actually separate from the Duhemian problem, is the problem that the qualitative comparative statics technique used in economics makes severe testing very difficult and cheap corroborational success 'too easy'. In economics it is very often the case that the strongest available prediction is a qualitative comparative statics result which only specifies that the variable in question increases, decreases, or remains the same. Since having the correct sign is much easier than having both the correct sign and magnitude, an emphasis on such qualitative prediction generates theories which are low in empirical content, have few potential falsifiers, and are difficult if not impossible to test severely. The result is often economic theories which are confirmed by the evidence but provide very little information.[10]
3. Popper's 'admitted failure' (1983: xxxv) to develop an adequate theory of verisimilitude[11] presents a fundamental difficulty for a falsificationist methodology in economics. Popper's theory of verisimilitude developed as an attempt to reconcile his falsificationist methodology with scientific realism. For a realist science aims at 'true' theories; according to falsificationism, scientific theories should be chosen if they have been corroborated by passing severe tests. If the falsificationist method is to fulfil the realist aims of science it should be demonstrated that more corroborated theories are closer to the truth. Such a demonstration was precisely the goal of Popper's theory of verisimilitude. Actually a satisfactory theory of verisimilitude would serve Popperian philosophy in at least two different ways. One way, as already mentioned, would be to provide an epistemic justification for playing the game of science by falsificationist rules. Such a justification is very important for Popperian philosophy since without a theory of verisimilitude it can be argued that there are philosophically 'no good reasons' (Popper 1972: 22) for choosing theories as Popper recommends. The second function of a theory of verisimilitude is more practical. Verisimilitude would provide rules for choosing the 'best' theory in troublesome cases: like the situation where both available theories have been falsified. A theory of verisimilitude would help in such cases because it would provide a rule for determining which of the two theories in question actually has more verisimilitude: which is closer to the truth. A similar argument could be made for cases involving a choice between a falsified but bold theory and a corroborated but modest theory; having a way to determine which has more verisimilitude would allow us to choose a theory which is more consistent with the aims of science, which is

closer to the truth. This second, more practical, function of the theory of verisimilitude is very important in economic methodology. The reason is that economists are almost always faced with choosing between two falsified theories, or choosing between a bold falsified theory and a more modest corroborated one. If Popper's theory of verisimilitude had been a success and it could be added to the norms of simple falsificationism (both to justify the norms and to help in making the practical decisions of theory choice) then falsificationism might have an important role to play in economic theory choice. Without such a link between severe testing and truth-likeness, the method is of limited value in pursuing the realist aim of science.

4. Popper's rules for progressive theory development (non *ad hocness*) are seldom appropriate in economics. Popper argues that if one theory is to constitute 'progress' over a predecessor the new theory must be 'independently testable'; it must have 'excess empirical content', predict 'novel facts'.[12] This issue will be examined more carefully in the Lakatos section which follows, but for now it should be noted that while Popperian progress may sometimes be of interest to economists, often progress in economics is (and should be) very different to what Popper prescribes. Economists are often concerned with finding new explanations for well-known (non novel) facts, or alternatively, with explaining known phenomena by means of fewer theoretical restrictions. What constitutes 'progress' in economic theory (or what should constitute progress) is a complex and ongoing question, but it is apparent that any suitable answer will require a different, and possibly much more liberal, set of standards than those offered by strict Popperian falsificationism.

All of these criticisms add up to a negative appraisal of falsificationist economic methodology. Despite the fact that preaching falsificationist methodology has been very popular among economists, the method fails to provide a reasonably adequate set of rules for doing economics. Strict adherence to falsificationist norms would virtually *destroy all existing economic theory* and leave economists with a rule book for a game unlike anything the profession has played in the past. This high cost would be paid without any guarantee that obeying the new rules would result in theories any closer to the truth about economic behaviour than those currently available. How this result should be interpreted will be discussed in the conclusion, for now let us turn to Lakatos's MSRP.

Lakatos's Methodology of Scientific Research Programmes

Lakatos's work in the philosophy of science first appeared in the early 1970s (Lakatos 1970, 1971) and it was endorsed almost immediately by a number of economists. Numerous papers on Lakatos have appeared in the economics literature, many as a result of the Nafplion Colloquium on Research Programmes in Physics and Economics in 1974 (Latsis 1976a). This literature on 'Lakatos and economics' has basically been of two types. The first type is historical, it attempts to 'reconstruct' some particular episode in the history of economic thought along Lakatosian lines. The second type is more philosophical, it attempts to appraise Lakatos's methodology of scientific research programmes as an economic methodology and/or compare it to other philosophies such as Kuhn or Popperian falsificationism.

Lakatos's MSRP is clearly part of the Popperian tradition in the philosophy of science but it was also motivated by philosophically minded historians of science such as Kuhn (1970). For Lakatos the primary unit of appraisal in science is the 'research programme' rather than the scientific theory. A research programme is an ensemble consisting of a hard core, the positive and negative heuristics, and a protective belt.[13] The hard core is composed of the fundamental metaphysical presuppositions of the programme; it defines the programme, and its elements are treated as irrefutable by the programme's practitioners. To participate in the programme is to accept and be guided by the programme's hard core. For example, in Weintraub's Lakatosian reconstruction of the Neo-Walrasian research programme in economics, the hard core consists of propositions such as: agents have preferences over outcomes and agents act independently and optimize subject to constraints. The positive and negative heuristics provide instructions about what should and should not be pursued in the development of the programme. The positive heuristic guides the researcher toward the right questions and the best tools to use in answering those questions; the negative heuristic indicates what questions should not be pursued and what tools are inappropriate. Again using Weintraub's analysis of the Neo-Walrasian programme as an example, the positive heuristic contains injunctions such as: construct theories where the agents optimize, while the negative heuristic implores researchers to avoid theories involving disequilibrium. Finally, the protective belt consists of the programme's actual theories, auxiliary hypotheses, empirical conventions and the (evolving) 'body' of the research programme. The major activity of the programme occurs in the protective belt, it occurs as a result of the interaction of the hard core, the heuristics, and the programme's empirical record. For Weintraub's Neo-Walrasian programme the protective belt includes almost all of applied microeconomics.

A research programme is appraised on the basis of the theoretical and empirical activity in the protective belt. There is *theoretical progress* if each change in the protective belt is empirical content increasing; that is if it predicts *novel facts*.[14] The research programme exhibits *empirical progress* if this excess empirical content actually gets corroborated (Lakatos 1970: 118). Lakatos also requires a third type of progress, heuristic progress (non-ad hoc$_3$ness), which specifies that the changes be consistent with the hard core of the programme. Lakatos's definitions of theoretical and empirical progress presuppose that the changes in question are consistent with heuristic progress.

One obvious example of the link between Lakatos and Popper is the way in which Lakatos characterizes empirical content and novel facts. Lakatos, like Popper, defines the empirical content of a theory to be 'the set of its potential falsifiers: the set of those observational propositions which may disprove it' (Lakatos 1970: 98, n. 2). Thus, even though Lakatos considers empirical progress to come through empirical corroboration rather than falsification, his characterization of the relationship between theory and fact is still basically falsificationist. There are many other signs of Lakatos's Popperian lineage but his definition of empirical content and novel facts are the most important in the appraisal of Lakatosian economic methodology.

On the other hand, there are many aspects of the MSRP which are fundamentally at odds with Popperian falsificationism. The most significant of these is the immunity of the hard core to empirical criticism; immunizing any part of scientific theory would be in conflict with Popper's falsificationist method of bold conjecture and severe test. Popper clearly recognized that science has experienced periods of Kuhnian 'normal science' where the critical spirit seems to be temporarily arrested, but for Popper these episodes are something to lament not praise (Popper 1970). Another point of disagreement is the question of corroboration versus falsification. While Lakatos defines empirical content in a thoroughly Popperian way, he has no respect for the role of falsification in science. For Lakatos all theories are 'born refuted' (1970: 120–1) and the task of philosophy of science should be to develop a methodology which *starts* from this fact. For Lakatos progress comes from the corroboration not falsification of novel facts. Finally, Lakatos clearly embraces a historical meta-methodology whereby the actual history of science is used to appraise various methodological proposals.[15] This is very different from Popper where methodology is purely a normative affair and where there is no pathway open for the actual history of science to help evaluate methodologies.

These places where Lakatos differs from Popper are exactly the places where Lakatos is likely to win the favour of economists since these happen to be areas where there is substantial tension between falsificationism and the actual practice of economics. Certainly economics is replete with metaphysical 'hard cores'; there is not much consensus on what these hard core propositions should be, but there seems to be a consensus that such hard core presuppositions exist and that they often define alternative research programmes in economics. A philosophical programme such as Popperian falsificationism which requires practitioners to be willing to give up almost any part of their research programme at any time will not provide as adequate a guide for economists as Lakatos's methodology which allows for such pervasive hard cores. This economic preference for Lakatos over Popper also extends to the issue of corroboration versus falsification. It is clear that falsificationism has not been practised in economics and there is good reason to believe that enforcement of such strict standards would all but eliminate the discipline as it currently exists. On the other hand, there *is* a great amount of empirical activity in economics, the facts do matter, but they matter in a much more subtle and complex way than falsificationism allows.

Finally, economists would prefer Lakatos to Popper on the question of the role of the history of science in supporting particular methodological proposals. The general question of the relationship between the history of science and the philosophy of science is an unsettled question which continues to be debated in the literature, but economists have recently been very sympathetic to methodological proposals that are sensitive to the actual history of their discipline. Economists have produced an extensive literature using the Lakatosian categories to reconstruct various parts of the history of economic thought. Most of this literature focuses on a particular research programme in economic theory (past or present) and tries to isolate the hard core, the positive and negative heuristics, and the type of theoretical activity occurring in the protective belt. Such work usually results in a positive or negative Lakatosian appraisal of the 'progressivity' of the particular economic research programme. Examples of these reconstructions range widely over various topics in the history of economic thought.

An overall assessment of this Lakatosian historical literature is very difficult because many of the economists writing in the field have taken very little care in the way they use the Lakatosian terminology. This lack of fidelity to Lakatosian terminology has resulted in 'hard cores', 'heuristics' and (particularly) 'novel facts' which bear little resemblance to their Lakatosian analogues or how these terms have been used in reconstructions in the physical sciences. Much of this literature has provided valuable and independently interesting history of economic thought, but it sheds little light on the methodological adequacy of the MSRP. The only general conclusion that can be reached from this historical literature is that *in the case studies where the relevant language is consistent with Lakatos, 'progress', and the prediction of novel facts it necessarily implies, has been a rare occurrence.* There have been some well-researched cases where novel facts actually seem to have been uncovered;[16] but these cases correspond to only a very small portion of what the economics profession would consider its major theoretical 'advances'. Lakatos's criterion for 'theoretical progress', the prediction of novel facts, may have been sufficient for what economists have considered to be theoretical progress in certain special cases, but it does not seem to be generally necessary. Just as 'the development of economic analysis would look a dismal affair through falsificationist spectacles' (Latsis 1976b: 8), it seems that economics would look almost as dismal on a strictly Lakatosian view.

The argument that empirical and theoretical advances in economics occur (and should occur) in ways other than Lakatos specified in the MSRP, reflects very poorly (again) on Popper. The reason is that *where economics is most likely to part ways with Lakatos is precisely where Lakatos borrowed most heavily from Popper*. In certain respects, Lakatos's work is much better suited to economics than Popper's; it seems that looking for the types of things which Lakatos suggests one should look for in the history of economics has helped guide a number of important historical studies. Certainly this historical research has drawn attention to the metaphysical hard core of certain economic research programmes and it has motivated enquiry into the important methodological question of the relationship between empirical and theoretical work in economics, that is, between econometrics and economic theory. What the MSRP does not provide is an appropriate model for the acceptance or rejection of economic theories. Lakatos's MSRP may constitute methodological progress over falsificationism, but it still fails to provide economists with an acceptable criterion for theory choice (or progressive problem shifts). This is particularly telling for Popper since the Lakatosian fit seems to be poorest where older Popperian parts were used with the least modification.

Conclusion

In the final evaluation it seems that 'Popperian' economic methodology must be given low marks. Falsificationism, Popper's fundamental programme for the growth of scientific knowledge, is particularly ill-suited to economics and while the interest in Lakatos has produced some valuable historical studies,[17] the overall fit between economics and the MSRP is not good: and not good precisely where Lakatos is the most Popperian.

This evaluation should not be too harshly interpreted though. It has been argued that Popperian methodology, both in its falsificationist and MSRP forms, does not provide a very good standard for judging the adequacy of economic theories; this does not mean that Popperian philosophy has not provided any insight at all into economic theorizing. In particular, the above argument does *not* say that testing should be unimportant in economics, that Lakatosian reconstructions in the history of economic thought have not provided valuable contributions to the historical literature, or that economists would have gained more by listening to some other particular school of philosophy.

In addition to the above disclaimers it should also be noted that the discussion has entirely neglected Popper's writings on the philosophy of *social science*: his so-called 'situational analysis' approach to social science.[18] This method, the method of explaining the behaviour of a social agent on the basis of the logic of the agent's situation and the 'rationality principle', was proposed by Popper as a result of 'the logical investigation of economics' and it provides a method 'which can be applied to all social science' (Popper 1976a: 102). It is often argued that the rationality principle is in conflict with Popper's falsificationist standards,[19] but regardless of how one views this controversy, the point here is simply to note that none of the above criticisms automatically transfer to Popper's work on situational analysis.

The task of this chapter was narrowly defined: to evaluate falsificationism and the MSRP as a methodology – as a tool for choosing between/among economic theories/research programmes. It has been argued that Popperian philosophy should be negatively appraised in this respect, it does not say that economists have nothing to learn from the Popperian tradition.

Notes

1. Helpful comments on an earlier draft were received from a number of people; in particular I would like to mention Bruce Caldwell, Christian Knudsen, Uskali Maki, and Jorma Sappinen. Partial support for the research was provided by University of Puget Sound Martin Nelson Award MNSA-4489 and portions of the argument also appear in Hands (1992). The recent article by Caldwell (1991) also provides an excellent discussion of these issues.
2. Blaug (1976, 1991), Cross (1982), de Marchi (1976), Diamond (1988), Fisher (1986), Fulton (1984), Glass and Johnson (1988), Hands (1985b), Latsis (1972, 1976b), Maddock (1984), and Weintraub (1985a,b, 1988), is a partial list of the

work on Lakatosian economics. A more complete list is contained in Hands (1985b) and (1992).

3. The expression 'basic statement' has a rather narrow interpretation in Popperian philosophy. The concept was introduced in chapter V of Popper (1959) and it is nicely summarized in Watkins (1984: 247–54).
4. Actually, as will be discussed below, scientific theories are not *by themselves* logically falsifiable. Rather, scientific theories along with (usually numerous) auxiliary hypotheses may form logically falsifiable *test systems* (see Hausman 1988: 68–9).
5. There are a number of different types of *ad hocness* in Popperian philosophy; these are discussed in detail in Hands (1988). The type of *ad hocness* considered here, modification solely to avoid falsification, is called *ad hoc*$_1$. Popper developed his notion of independent testability to avoid this type of *ad hocness (ad hoc*$_1$*ness)*. Another notion of *ad hocness* is *ad hoc*$_2$*ness*; a theoretical modification is non *ad hoc*$_2$ if some of the independently testable implications actually get corroborated. A third type of *ad hocness (ad hoc*$_3$*ness)* was developed more fully by Lakatos. *Non-ad hoc*$_3$*ness* is equivalent to Lakatosian heuristic progress.
6. The main sources for this list of criticisms are Caldwell (1984), Hausman (1985, 1988), Latsis (1976b), and Salanti (1987).
7. The Duhemian problem (Duhem 1954) arises because theories are never tested alone, rather they are tested in conjunction with certain auxiliary hypotheses (including those about the data). Thus if T is the theory, the prediction of evidence e is given by $T \cdot \mathrm{A} \Rightarrow \mathrm{e}$, where A is the set of auxiliary hypotheses. The conjunction $T \cdot \mathrm{A}$ forms a test system and the observation of 'not *e*' implies 'not $(T \cdot \mathrm{A})$' rather than simply 'not T'; the test system is falsified, not necessarily the theory. The Duhemian problem is a standard issue in the philosophy of theory testing but it has only recently been recognized as an issue for economic methodology (see Cross 1982, for instance).
8. Experimental economics is still too young to tell whether it can substantially improve this situation. For a general discussion of the methodological implications of the literature on experimental economics see Roth (1986), and Smith (1982, 1985).
9. Popper (1965: 42, 267, 387–8; 1959: 43–4, 93–5, 97–111; 1983: 185–6).
10. This is one source of the 'innocuous falsification' mentioned by Blaug (1980: 128, 259) and Coddington (1975: 542–45). The problem of such qualitative (or generic) predictions is discussed in detail in Rosenberg (1989).
11. Popper's most important writings on verisimilitude are contained in Popper (1965) and (1972). Useful discussions of the topic are presented in Koertge (1979), Radnitzky (1982), and Watkins (1984). The question of the relationship between Popperian verisimilitude and economic methodology is examined in more detail in Hands (1991).
12. These concepts are discussed in detail with appropriate references to Popper's writings in Hands (1988). Other general discussions of these Popperian concepts include Dilworth (1986), Koertge (1979), Watkins (1978, 1984), and Worrall (1978).

13. Many summaries of the MSRP are available in the economics literature (Blaug (1980), Hands (1985a), Pheby (1988), and Weintraub (1985a, 1985b, 1988) for example) but the single best presentation of the argument remains Lakatos (1970). As with Popper's falsificationism, only a sketch of the main argument is provided here.
14. The definition of 'novel fact' has been much discussed in the Lakatosian (and Popperian) literature. See Carrier (1988), Gardner (1982), Hands (1985b) and Worrall (1978) on the different notions of novelty.
15. 'A general definition of science thus must reconstruct the acknowledgedly best gambits as "scientific:" if it fails to do so, it has to be rejected' (Lakatos 1971: 111).
16. Particularly Blaug (1991), Maddock (1984), and Weintraub (1988), though even here it depends on the exact definition of novelty one uses.
17. In addition to those mentioned in note 16, Cross (1982), de Marchi (1976) and Latsis (1972, 1976b) should be added to this list.
18. Popper's clearest writings on situational analysis are (1976a) and (1985); also see Hands (1985a, 1992) and Langlois and Csontos (this volume).
19. According to Popper's situational analysis view of social science, the action of an individual agent is explained by describing the 'situation' the agent is in (their preferences, beliefs, constraints, etc.) and the 'rationality principle' that all agents act appropriately (rationally) given their situation. The potential problem arises because the rationality principle serves as the universal law in such scientific 'explanations' and yet it is not clear that the rationality principle is (even potentially) falsifiable as Popper the falsificationist would require for the laws used in any valid scientific explanation. This is one of the reasons that Popper_n (Popper the falsificationist) was distinguished from Popper_s (Popper the philosopher of situational analysis) in Hands (1985a).

References

Blaug, M. (1976) 'Kuhn versus Lakatos, or paradigms versus research programmes in the history of economics', in S. J. Latsis (ed.) *Method and Appraisal in Economics*, Cambridge: Cambridge University Press, pp. 149–80.

——— (1980) *The Methodology of Economics*, Cambridge: Cambridge University Press.

——— (1987) 'Ripensamenti sulla rivduzione Keynesiana', *Rassegna Economica*, 51: 605–34.

——— (1991) 'Second thoughts on the Keynesian revolution', *History of Political Economy* 23: 171–92 (English translation of Blaug 1987).

Caldwell, B. J. (1982) *Beyond Positivism: Economic Methodology in the Twentieth Century*, London: Allen & Unwin.

——— (1984) 'Some problems with falsificationism in economics', *Philosophy of the Social Sciences* 14: 489–95.

——— (1991) 'Clarifying Popper', *Journal of Economic Literature* 29: 1–33.

Carrier, M. (1988) 'On novel facts: a discussion of criteria for non-ad-hocness in the methodology of scientific research programmes', *Zeitschrift fur Allgemeine Wissenschaftstheorie* 19: 205–31.

Coddington, A. (1975) 'The rationale of general equilibrium theory', *Economic Inquiry* 13: 539–58.

Cross, R. (1982) 'The Duhem–Quine thesis, Lakatos, and the appraisal of theories in macroeconomics', *Economic Journal* 92: 320–40.

de Marchi, N. (1976) 'Anomaly and the development of economics: the case of the Leontief paradox', in S. J. Latsis (ed.) *Method and Appraisal in Economics*, Cambridge: Cambridge University Press, pp. 109–27.

——— (ed.) (1988) *The Popperian Legacy in Economics*, Cambridge: Cambridge University Press.

Diamond, A. M., Jr. (1988) 'The empirical progressiveness of the general equilibrium research program', *History of Political Economy* 20: 119–35.

Dilworth, C. (1986) *Scientific Progress: A Study Concerning the Nature of the Relations Between Successive Scientific Theories*, second edition, Dordrecht, Netherlands: D. Reidel.

Duhem, P. (1954) *The Aim and Structure of Physical Theory*, (translated by P. P. Wiener), Princeton, New Jersey: Princeton University Press.

Fisher, F. M. (1986) *The Logic of Economic Discovery*, Washington Square, New York: New York University Press.

Fulton, G. (1984) 'Research programmes in economics', *History of Political Economy* 16: 187–205.

Gardner, M. R. (1982) 'Predicting novel facts', *British Journal for the Philosophy of Science* 33: 1–15.

Gellner, E. (1974) *Legitimation of Belief*, Cambridge: Cambridge University Press.

Glass, J. C. and Johnson, W. (1988) 'Metaphysics, MSRP and economics', *British Journal for the Philosophy of Science* 39: 313–29.

Hands, D. W. (1985a) 'Karl Popper and economic methodology: *A New Look*', *Economics and Philosophy* 1: 83–99.

——— (1985b) 'Second thoughts on Lakatos', *History of Political Economy* 17: 1–16.

——— (1988) 'Ad hocness in economics and the Popperian tradition', in de Marchi, N. (ed.) *The Popperian Legacy in Economics and Beyond*, Cambridge: Cambridge University Press, pp. 121–37.

——— (1991) 'The problem of excess content: economics, novelty and a long Popperian Tale', in N. de Marchi and M. Blaug (eds) *Appraising Economic Theories*, Aldershot: Edward Elgar 58–75.

——— (1992) 'Falsification, situational analysis and scientific research programs: the Popperian tradition in economic methodology', in N. de Marchi (ed.) *Post-Popperian Methodology of Economics*, Boston: Kluwer.

Hausman, D. M. (1985) 'Is falsificationism unpracticed or unpracticable?', *Philosophy of the Social Sciences* 15: 313–19.

——— (1988) 'An appraisal of Popperian economic methodology', in de Marchi, N. (ed.) *The Popperian Legacy in Economics and Beyond*, Cambridge: Cambridge University Press, 65–85.

Hutchison, T. W. (1938) *The Significance and Basic Postulates of Economic Theory*, London: Macmillan (reprinted, 1960, New York: Augustus M. Kelley).

——— (1988) 'The case for falsificationism', in de Marchi, N. (ed.) *The Popperian Legacy in Economics and Beyond*, Cambridge: Cambridge University Press, pp. 169–81.

Koertge, N. (1979) 'The problems of appraising scientific theories', in P. D. Asquith and H. E. Kyburg, Jr. (eds) *Current Research in Philosophy of Science, East Lansing*, Michigan: Philosophy of Science Association, pp. 228–251.

Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*, second edition, Chicago: University of Chicago Press. (First edition 1962, Chicago: University of Chicago Press.)

Lakatos, I. (1970) 'Falsification and the methodology of scientific research programmes', in I. Lakatos and A. Musgrave (eds) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, pp. 91–196.

——— (1971) 'History of science and its rational reconstructions', in R. C. Buck and R. S. Cohen (eds) *Boston Studies in the Philosophy of Science*, Vol. 8, Dordrecht, Netherlands: D. Reidel, pp. 91–136.

Latsis, S. J. (1972) 'Situational determinism in economics', *British Journal for the Philosophy of Science* 23: 207–45.

——— (ed.) (1976a) *Method and Appraisal in Economics*, Cambridge: Cambridge University Press.

——— (1976b) 'A research programme in economics', in S. J. Latsis (ed.) *Method of Appraisal in Economics*, Cambridge: Cambridge University Press, pp. 1–41.

McCloskey, D. (1985) *The Rhetoric of Economics*, Madison Wisconsin: University of Wisconsin Press.

Maddock, R. (1984) 'Rational expectations macrotheory: A Lakatosian case study in program adjustment', *History of Political Economy* 16: 291–309.

Pheby, J. (1988) *Methodology and Economics: A Critical Introduction*, London: Macmillan.

Popper, K. R. (1959) *The Logic of Scientific Discovery*, London: Hutchinson.

——— (1965) *Conjectures and Refutations*, second edition, New York: Harper & Row.

——— (1970) 'Normal science and its Dangers', in I. Lakatos and A. Musgrave (eds) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, pp. 51–8.

——— (1972) *Objective Knowledge: An Evolutionary Approach*, Oxford: Oxford University Press.

——— (1976a) 'The logic of the social sciences', in T. W. Adorno et al. (eds) *The Positivist Dispute in German Sociology* (translated by G. Adey and D. Frisby), New York: Harper & Row, pp. 87–104.

——— (1976b) *Unended Quest*, La Salle, Illinois: Open Court.

——— (1983) *Realism and the Aim of Science* (edited by W. W. Bartley III) Totowa, New Jersey: Rowman & Littlefield.

——— (1985) 'The rationality principle', in D. Miller (ed.) *Popper Selections*, Princeton: Princeton Universtiy Press, pp. 357–65.

Radnitzky, G. (1982) 'Knowing and guessing: if all knowledge is conjectural, can we then speak of cognitive progress? on persistent misreading of Popper's work', *Zeitschrift für Allgemeine Wissenschafstheorie* 8: 110–21.

Rosenberg, A. (1989) 'Are generic predictions enough?', *Erkenntnis* 30: 43–68.

Roth, A. E. (1986) 'Laboratory experimentation in economics', *Economics and Philosophy* 2: 245–73.

Salanti, A. (1987) 'Falsificationism and fallibilism as epistemic foundations of economics: a critical view', *Kyklos* 40: 368–92.

Smith, V. L. (1982) 'Microeconomic systems as an experimental science', *American Economic Review* 72: 923–55.

——— (1985) 'Experimental economics: reply', *American Economic Review* 75: 265–72.

Watkins, J. (1978) 'The Popperian approach to scientific knowledge', in G. Radnitzky and G. Anderson (eds) *Progress and Rationality in Science*, Dordrecht, Netherlands: D. Reidel, pp. 23–43.

——— (1984) *Science and Skepticism*, Princeton, New Jersey: Princeton University Press.

Weintraub, E. R. (1985a) 'Appraising general equilibrium analysis', *Economics and Philosophy* 1: 23–37.

——— (1985b) *General Equilibrium Analysis: Studies in Appraisal*, Cambridge: Cambridge University Press.

——— (1988) 'The NeoWalrasian program is empirically progressive', in de Marchi, N. (ed.) *The Popperian Legacy in Economics and Beyond*, Cambridge: Cambridge University Press.

Worrall, J. (1978) 'The ways in which the methodology of scientific research programmes improves on Popper's methodology', in G. Radnitzky and G. Anderson (eds) *Progress and Rationality in Science*, Dordrecht, Netherlands: D. Reidel, pp. 45–70.

PART THREE

IDEOLOGY AND NORMATIVE ECONOMICS

Economic questions are often policy questions, and the answers affect the well-being of millions. Although many sciences have technical uses, the practical importance of economics is special. Indeed, in addition to "positive economics," which aims to predict and perhaps to explain economic phenomena, there is also "normative economics," which aims to guide policy. In addition, economics is built around a theory of rationality, which seems also to be normative, though in a different way.

The connections between economics and values raise several different kinds of methodological questions, which are discussed in the five essays in this Part III. In the first essay, Joseph Schumpeter explores how general evaluative visions as well as specific value commitments have shaped the development of economics. The next three essays by Nicholas Kaldor, Michael S. McPherson and myself, and Robert H. Frank lay out the foundations of normative economics and explore some of the difficulties to which it is subject. Finally, Amartya Sen's "Capability and Well-Being," presents a radical alternative to standard normative economics, which is currently attracting a great deal of attention.

ELEVEN

Science and Ideology

Joseph Schumpeter

Joseph Schumpeter (1883–1950) was born in Trietsch, Austria, and studied law and economics at the University of Vienna. He taught in Austria and Germany before coming to Harvard in the 1930s. The essay reprinted here was Schumpeter's presidential address to the American Economic Association in 1948. Schumpeter made major contributions to the understanding of economic growth and crises, and his *History of Economic Analysis* is perhaps the greatest work ever written on the history of economics. Schumpeter's *Capitalism, Socialism and Democracy* is also a major contribution to economics and political theory.

I

A hundred years ago economists were much more pleased with their performance than they are today. But I submit that, if complacency can ever be justified, there is much more reason for being complacent today than there was then or even a quarter of a century ago. As regards command of facts, both statistical and historical, this is so obviously true that I need not insist. And if it be true of our command of facts, it must be true also for all the applied fields that for their advance mainly depend upon fact finding. I must insist, however, on the proposition that our powers of analysis have grown in step with our stock of facts. A new organon of statistical methods has emerged, to some extent by our own efforts, that is bound to mean as much to us as it does to all the sciences, such as biology or experimental psychology, the phenomena of which are given in terms of frequency distributions. In response to this development and in alliance with it, as well as independently, our own box of analytic tools has been greatly enriched: economic theory, in the instrumental sense of the term – in which it means

Reprinted by permission of the American Economic Association from *American Economic Review*, vol. 39(1949): 345–59.

neither the teaching of ultimate ends of policy nor explanatory hypotheses but simply the sum total of our methods of handling facts – has grown quite as much as Marshall and Pareto had foreseen that it would.

If this is not more generally recognized and if it is etiquette with economists – let alone the public – to pass derogatory judgment on the state of our science, this is owing to a number of causes that, though known all too well, should be repeated: a building plot on which old structures are being torn down and new ones erected is not an esthetic thing to behold; moreover, to a most discouraging extent the new structures are being currently discredited by premature attempts at utilitarian application; finally, the building area widens so that it becomes impossible for the individual worker to understand everything that is going on beyond his own small sector. It would indeed be difficult to present in systematic form, as the Smiths, Mills, and Marshalls have been able to do with more or less success, a comprehensive treatise that might display some measure of unity and command all but universal approval. Thus, though the workers in each sector are not at all displeased with how they are getting on themselves, they are quite likely to disapprove of the manner in which those in all the others go about their tasks, or even to deny that these other tasks are worth bothering about at all. This is but natural. Many types of mind are needed to build up the structure of human knowledge, types which never quite understand one another. Science is technique and the more it develops, the more completely does it pass out of the range of comprehension not only of the public but, minus his own chosen specialty, of the research worker himself. More or less, this is so everywhere although greater uniformity of training and greater discipline of endeavor may in physics reduce the tumult to something like order. As everyone knows, however, there is with us another source of confusion and another barrier to advance: most of us, not content with their scientific task, yield to the call of public duty and to their desire to serve their country and their age, and in doing so bring into their work their individual schemes of values and all their policies and politics – the whole of their moral personalities up to their spiritual ambitions.

I am not going to reopen the old discussion on value judgments or about the advocacy of group interests. On the contrary, it is essential for my purpose to emphasize that *in itself* scientific performance does not require us to divest ourselves of our value judgments or to renounce the calling of an advocate of some particular interest. To investigate facts or to develop tools for doing so is one thing; to evaluate them from some moral or cultural standpoint is, *in logic*, another thing, and the two *need* not conflict. Similarly, the advocate of some interest may yet do honest analytic work, and the motive of proving a point for the interest to which he owes allegiance does not in itself prove anything for or against this analytic work: more bluntly, advocacy does not imply lying. It spells indeed misconduct to bend either facts or inferences from facts in order to make them serve either an ideal or an interest. But such misconduct is not necessarily inherent in a worker's arguing from "axiological premises" or in advocacy *per se*.[1] Examples abound in which economists have established propositions for the implications of which they did not have any sympathy. To mention a single instance: to establish the logical consistency of the conditions (equations) that are descriptive of a socialist economy will seem to most people equivalent to gaining a point for socialism; but it was established by Enrico Barone, a man who, whatever else he may have been, was certainly no sympathizer with socialist ideals or groups.

But there exist in our minds preconceptions about the economic process that are much more dangerous to the cumulative growth of our knowledge and the scientific character of our analytic endeavors because they seem beyond our control in a sense in which value judgments and special pleadings are not. Though mostly allied with these, they deserve to be separated from them and to be discussed independently. We shall call them Ideologies.

II

The word *idéologie* was current in France toward the end of the eighteenth and in the first decade of the nineteenth century and meant much the same thing as did the Scottish moral philosophy of the same and an earlier time or as our own social science in that widest acceptance of the term in which it includes psychology. Napoleon imparted a derogatory meaning to it by his sneers at the *idéologues* – doctrinaire dreamers without any sense for the realities of politics. Later on, it was used as it is often used today in order to denote systems of ideas, that is, in a way in which our distinction between ideologies and value judgments is lost. We have nothing to do with these or any other meanings except one that may be most readily introduced by reference to the "historical materialism" of Marx and Engels. According to this doctrine, history is determined by the autonomous evolution of the structure of production: the social and political organization, religions, morals, arts and sciences are mere "ideological superstructures," generated by the economic process.

We neither need nor can go into the merits and demerits of this conception as such[2] of which only one feature is relevant to our purpose. This feature is the one that has, through various transformations, developed into the sociology of science of the type associated with the names of Max Scheler and Karl Mannheim. Roughly up to the middle of the 19th century the evolution of "science" had been looked upon as a purely intellectual process – as a sequence of explorations of the empirically given universe or, as we may also put it, as a process of filiation of discoveries or analytic ideas that went on, though no doubt influencing social history and being influenced by it in many ways, according to a law of its own. Marx was the first to turn this relation of interdependence between "science" and other departments of social history into a relation of dependence of the former on the objective data of the social structure and in particular on the social location of scientific workers that determines their outlook upon reality and hence what they see of it and how they see it. This kind of relativism – which must of course not be confused with any other kind of relativism[3] – if rigorously carried to its logical consequences spells a new philosophy of science and a new definition of scientific truth. Even for mathematics and logic and still more for physics, the scientific worker's choice of problems and of approaches to them, hence the pattern of an epoch's scientific thought, becomes socially conditioned – which is precisely what we mean when speaking of scientific ideology rather than of the ever more perfect perception of objective scientific truths.

Few will deny, however, that in the cases of logic, mathematics, and physics the influence of ideological bias does not extend beyond that choice of problems and approaches, that is to say, that the sociological interpretation does not, at least for the last two or three centuries, challenge the "objective truth" of the findings. This "objective truth" may be, and currently is being, challenged on other grounds but not on the ground that a given proposition is true only with reference to the social location of the men who formulated it. To some extent at least, this favorable situation may be accounted for by the fact that logic, mathematics, physics and so on deal with experience that is largely invariant to the observer's social location and practically invariant to historical change: for capitalist and proletarian, a falling stone looks alike. The social sciences do not share this advantage. It is possible, or so it seems, to challenge their findings not only on all the grounds on which the propositions of all sciences may be challenged but also on the additional one that they cannot convey more than a writer's class affiliations and that, without reference to such class affiliations, there is no room for the categories of true or false, hence for the conception of "scientific advance" at all. Henceforth we adopt the term Ideology or Ideological Bias for this – real or supposed – state of things alone, and our problem is to ascertain the extent to which ideological bias is or has been a factor in the development of what – conceivably – it might be a misnomer to call scientific economics.

In recognizing the ideological element it is possible to go to very different lengths. There are a few writers who have in fact denied that there is such a thing in economics as accumulation of a stock of "correctly" observed facts and "true" propositions. But equally small is the minority who would deny the influence of ideological bias entirely. The majority of economists stand between these extremes: they are ready enough to admit its presence though, like Marx, they find it only in others and never in themselves; but they do not admit that it is an inescapable curse and that it vitiates economics to its core. It is precisely this intermediate position that raises our problem. *For ideologies are not simply lies*; they are truthful statements about what a man thinks he sees. Just as the medieval knight saw himself as he wished to see himself and just as the modern bureaucrat does the same and just as both failed and fail to see whatever may be adduced against their seeing themselves as the defenders of the weak and innocent and the sponsors of the Common Good, so every other social group develops a protective ideology which is nothing if not sincere. *Ex hypothesi* we are not aware of our rationalizations – how then is it possible to recognize and to guard against them?

But let me repeat before I go on: I am speaking of science which is technique that turns out the results which, together with value judgments or preferences, produce recommendations, either individual ones or systems of them – such as the systems of mercantilism, liberalism, and so on. I am not speaking of these value judgments and these recommendations themselves. I fully agree with those who maintain that judgments about ultimate values – about the Common Good, for instance – are beyond the scientist's range except as objects of historical study, that they are ideologies by nature and that the concept of scientific progress can be applied to them only so far as the means may be perfected that are to implement them. I share the conviction that there is no sense in saying that the world of ideas of bourgeois liberalism is "superior" in any relevant sense to the world of ideas of the middle ages, or the world of ideas of socialism to that of bourgeois liberalism. Actually, I further believe that there is no reason other than personal preference for saying that more wisdom or knowledge goes into our policies than went into those of the Tudors or Stuarts or, for that matter, into Charlemagne's.

III

So soon as we have realized the possibility of ideological bias, it is not difficult to locate it. All we have to do for this purpose is to scrutinize scientific procedure. It starts from the perception of a set of related phenomena which we wish to analyze and ends up – for the time being – with a scientific model in which these phenomena are conceptualized and the relations between them explicitly formulated, either as assumptions or as propositions (theorems). This primitive way of putting it may not satisfy the logician but it is all we need for our hunt for ideological bias. Two things should be observed.

First, that perception of a set of related phenomena is a prescientific act. It must be performed in order to give to our minds something to do scientific work on – to indicate an object of research – but it is not scientific in itself. But though prescientific, it is not preanalytic. It does not simply consist in perceiving facts by one or more of our senses. These facts must be recognized as having some meaning or relevance that justifies our interest in them and they must be recognized as related – so that we might separate them from others – which involves some analytic work by our fancy or common sense. This mixture of perceptions and prescientific analysis we shall call the research worker's Vision or Intuition. In practice, of course, we hardly ever start from scratch so that the prescientific act of vision is not entirely our own. We start from the work of our predecessors or contemporaries or else from the ideas that float around us in the public mind. In this case our vision will also contain at least some of the results of previous scientific analysis. However, this compound is still given to us and exists before we start scientific work ourselves.

Second, if I have identified with "model building" the scientific analysis that operates upon the material proffered by the vision, I must add at once that I intend to give the term "model" a very wide meaning. The explicit economic model of our own day and its analoga in other sciences are of course the product of late stages of scientific endeavor. Essentially, however, they do not do anything that is not present in the earliest forms of analytic endeavor which may therefore also be said to have issued, with every individual worker, in primitive, fragmentary, and inefficient models. This work consists in picking out certain facts rather than others, in pinning them down by labeling them, in accumulating further facts in order not only to supplement but in part also to replace those originally fastened upon, in formulating and improving the relations perceived – briefly, in "factual" and "theoretical" research that go on in an endless chain of give and take, the facts suggesting new analytic instruments (theories) and these in turn carrying us toward the recognition of new facts. This is as true when the object of our interest is an historical report as it is when the object of our interest is to "rationalize" the Schrödinger equation though in any particular instance the task of fact finding or the task of analyzing may so dominate the other as to almost remove it from sight. Schoolmasters may try to make this clearer to their pupils by talking about induction and deduction and even set the one against the other, creating spurious problems thereby. The essential thing, however we may choose to interpret it, is the "endless give and take" between the clear concept and the cogent conclusion on the one hand, and the new fact and the handling of its variability on the other.

Now, so soon as we have performed the miracle of knowing what we cannot know, namely the existence of the ideological bias in ourselves and others, we can trace it to a simple source. This source is in the initial vision of the phenomena we propose to subject to scientific treatment. For this treatment itself is under objective control in the sense that it is always possible to establish whether a given statement, in reference to a given state of knowledge, is provable, refutable, or neither. Of course this does not exclude honest error or dishonest faking. It does not exclude delusions of a wide variety of types. But it does permit the exclusion of that particular kind of delusion which we call ideology because the test involved is indifferent to any ideology. The original vision, on the other hand, is under no such control. There, the elements that will meet the tests of analysis are, by definition, undistinguishable from those that will not or – as we may also put it since we admit that ideologies *may* contain provable truth up to 100 percent – the original vision *is* ideology by nature and may contain any amount of delusions traceable to a man's social location, to the manner in which he wants to see himself or his class or group and the opponents of his own class or group. This should be extended even to peculiarities of his outlook that are related to his personal tastes and conditions and have no group connotation – there is even an ideology of the mathematical mind as well as an ideology of the mind that is allergic to mathematics.

It may be useful to reformulate our problem before we discuss examples. Since the source of ideology is our pre- and extrascientific vision of the economic process and of what is – causally or teleologically – important in it and since normally this vision is then subjected to scientific treatment, it is being either verified or destroyed by analysis and in either case should vanish *qua* ideology. How far, then, does it fail to disappear as it should? How far does it hold its own in the face of accumulating adverse evidence? And how far does it vitiate our analytic procedure itself so that, in the result, we are still left with knowledge that is impaired by it?

From the outset it is clear that there is a vast expanse of ground on which there should be as little danger of ideological vitiation as there is in physics. A time series of gross investment in manufacturing industry may be good or bad, but whether it is the one or the other is, normally, open to anyone to find out. The Walrasian system *as it stands* may or may not admit of a unique set of solutions but whether it does or not is a matter of exact proof that every qualified person can repeat. Questions like these may not be the most fascinating or practically most urgent ones but they constitute the bulk of what is specifically scientific in our work. And they are in logic although not always in fact neutral to ideology. Moreover, their sphere widens as our understanding of analytic work improves. Time was when economists thought that they were gaining or losing a point for labor if they fought for the labor-quantity and against the marginal-utility theory of value. It can be shown that, so far as ideologically relevant issues are concerned, this makes as little difference as did the replacement of the latter by the indifference-curve approach or the replacement of the indifference curves by a simple consistency postulate (Samuelson). I dare say that there are still some who find something incongruous to their vision in marginal-productivity analysis. Yet it can be shown that the latter's purely formal apparatus is compatible with any vision of economic reality that anyone ever had.[4]

IV

Let us now look for ideological elements in three of the most influential structures of economic thought, the works of Adam Smith, of Marx, and of Keynes.

In Adam Smith's case the interesting thing is not indeed the absence but the harmlessness of ideological bias. I am not referring to his time- and country-bound practical wisdom about *laissez-faire*, free trade, colonies and the like for – it cannot be repeated too often – a man's political preferences and recommendations as such are entirely beyond the range of my remarks or rather they enter this range only so far as the factual and theoretical analysis does that is presented in support of them. I am exclusively referring to this analytical work itself – only to his indicatives, not to his imperatives. This being understood, the first question that arises is what kind of ideology we are to attribute to him. Proceeding on the Marxist principle we shall look to his social location, that is, to his personal and ancestral class affiliations and in addition to the class connotation of the influences that may have formed or may have helped to form what we have called his vision. He was a *homo academicus* who became a civil servant. His people were more or less of a similar type: his family, not penniless but neither wealthy, kept up some standard of education and fell in with a well-known group in the Scotland of his day. Above all it did *not* belong to the business class. His general outlook on things social and economic reproduced these data to perfection. He beheld the economic process of his time with a cold critical eye and instinctively looked for mechanical rather than personal factors of explanation – such as division of labor. His attitude to the land-owning and to the capitalist classes was the attitude of the observer from outside and he made it pretty clear that he considered the landlord (the "slothful" landlord who reaps where he has not sown) as an unnecessary, and the capitalist (who hires "industrious people" and provides them with subsistence, raw materials, and tools) as a necessary evil. The latter necessity was rooted in the virtue of parsimony, eulogy of which evidently came from the bottom of his Scottish soul. Apart from this, his sympathies went wholly to the laborer who "clothes everybody and himself goes in rags." Add to this the disgust he felt – like all the people in his group – at the inefficiency of the English bureaucracy and at the corruption of the politicians and you have practically all of his ideological vision. While I cannot stay to show how much this explains of the picture he drew, I must emphasize that the other component of this vision, the natural-law philosophy that he imbibed in his formative years, the product of similarly conditioned men, influenced the ideological background from which he wrote in a similar manner – natural freedom of action, the workman's natural right to the whole product of industry, individualistic rationalism and so on, all this was taught to him ere his critical faculties were developed but there was hardly need to teach him these things for they came "naturally" to him in the air he breathed. But – and this is the really interesting point – all this ideology, however strongly held, really did not do much harm to his scientific achievement. Unless we go to him for economic sociology,[5] we receive from him sound factual and analytic teaching that no doubt carries date but is not open to objection on the score of ideological bias. There is some semiphilosophical foliage of an ideological nature but it can be removed without injury to his scientific argument. The analysis that supports his qualified free-trade conclusions is not – as it was with some contemporaneous philosophers, such as Morellet – based upon the proposition that by nature a man is free to buy or to sell where he pleases. The statement that the (whole) produce is the natural compensation of labor occurs, but no analytic use is made of it – everywhere the ideology spends itself in phraseology and for the rest recedes before scientific research. In part at least, this was the merit of the man: he was nothing if not responsible; and his sober and perhaps somewhat dry common sense gave him respect for facts and logic. In part it was good fortune: it matters little if his analysis has to be given up as the *psychology* it was meant to be if at the same time it must be retained as a *logical* schema of economic behavior – on closer acquaintance, the *homo economicus* (so far as Adam Smith, the author of the *Moral Sentiments*, can in fact be credited or debited with this conception at all) turns out to be a very harmless man of straw.

Marx was the economist who discovered ideology for us and who understood its nature. Fifty years before Freud, this was a performance of the first order. But, strange to relate, he was entirely blind to its dangers so far as he himself was concerned. Only other people, the bourgeois economists and the utopian socialists, were victims of ideology. At the same time, the ideological character of his premises and the ideological bias of his argument are everywhere obvious. Even some of his followers (Mehring, for instance) recognized this. And it is not difficult to describe his ideology. He was a bourgeois radical who had broken away from bourgeois radicalism. He was formed by German philosophy and did not feel himself to be a professional economist until the end of the 1840's. But by that time, that is to say, *before* his serious analytic work had begun, his vision of the capitalist process had become set and his scientific work was to implement, not to correct it. It was not original with him. It pervaded the radical circles of Paris and may be traced back to a number of 18th century writers, such as Linguet.[6] History conceived as the struggle between classes that are defined as *haves* and *havenots*, with exploitation of the one by the other, ever increasing wealth among even fewer *haves* and ever increasing misery and degradation among the *havenots*, moving with inexorable necessity toward spectacular explosion, this was the vision then conceived with passionate energy and to be worked up, like a raw material is being worked up, by means of the scientific tools of his time. This vision implies a number of statements that will not stand the test of analytic controls. And, in fact, as his analytic work matured, Marx not only elaborated many pieces of scientific analysis that were neutral to that vision but also some that did not agree with it well – for instance, he got over the kind of underconsumption and the kind of overproduction theories of crises which he seems to have accepted at first and traces of which – to puzzle interpreters – remained in his writings throughout. Other results of his analysis he introduced by means of the device of retaining the original – ideological – statement as an "absolute" (*i.e.*, abstract) law while admitting the existence of counteracting forces which accounted for deviating phenomena in real life. Some parts of the vision, finally, took refuge in vituperative phraseology that does not affect the scientific elements in an argument. For instance, whether right or wrong, his exploitation theory of "surplus" value was a genuine piece of theoretical analysis. But all the glowing phrases about exploitation could have been attached just as well to other theories, Böhm-Bawerk's among them: imagine Böhm-Bawerk in

Marx's skin, what could have been easier for him than to pour out the vials of his wrath on the infernal practice of robbing labor by means of deducting from its product a time discount?

But some elements of his original vision – in particular the increasing misery of the masses which was what was to goad them into the final revolution – that were untenable were at the same time indispensable for him. They were too closely linked to the innermost meaning of his message, too deeply rooted in the very meaning of his life, to be ever discarded. Moreover, they were what appealed to followers and what called forth their fervent allegiance. It was they which explain the organizing effect – the party-creating effect – of what without them would have been stale and lifeless. And so we behold in this case the victory of ideology over analysis: all the consequences of a vision that turns into a social creed and thereby renders analysis sterile.

Keynes' vision – the source of all that has been and is more or less definitely identified as Keynesianism – appeared first in a few thoughtful paragraphs in the introduction to the *Consequences of the Peace* (1920). These paragraphs created *modern* stagnationism – stagnationist moods had been voiced, at intervals, by many economists before, from *Britannia Languens* on (1680) – and indicate its essential features, the features of mature and arteriosclerotic capitalist society that tries to save more than its declining opportunities for investment can absorb. This vision never vanished again – we get another glimpse of it in the tract on *Monetary Reform* and elsewhere but, other problems absorbing Keynes' attention during the 1920's, it was not implemented analytically until much later. D. H. Robertson in his *Banking Policy and the Price Level* presented some work that amounted to partial implementation of the idea of abortive saving. But with Keynes this idea remained a side issue even in the *Treatise on Money*. Perhaps it was the shock imparted by the world crisis which definitely broke the bonds that prevented him from fully verbalizing himself. Certainly it was the shock imparted by the world crisis which created the public for a message of this kind.

Again it was the ideology – the vision of decaying capitalism that located (*saw*) the cause of the decay in one out of a large number of features of latter-day society – which appealed and won the day, and not the analytic implementation by the book of 1936 which, by itself and without the protection it found in the wide appeal of the ideology, would have suffered much more from the criticisms that were directed against it almost at once. Still, the conceptual apparatus was the work not only of a brilliant but also of a mature mind – of a Marshallian who was one of the three men who had shared the sage's mantle between them. Throughout the 1920's Keynes was and felt himself to be a Marshallian and even though he later on renounced his allegiance dramatically, he never deviated from the Marshallian line more than was strictly necessary in order to make his point. He continued to be what he had become by 1914, a master of the theorist's craft, and he was thus able to provide his vision with an armour that prevented many of his followers from seeing the ideological element at all. Of course this now expedites the absorption of Keynes' contribution into the current stream of analytic work. There are no really new principles to absorb. The ideology of underemployment equilibrium and of non-spending – which is a better term to use than saving – is readily seen to be embodied in a few restrictive assumptions that emphasize certain (real or supposed) facts. With these everyone can deal as he thinks fit and for the rest he can continue his way. This reduces Keynesian controversies to the level of technical science. Lacking institutional support, the "creed" has petered out with the situation that had made it convincing. Even the most stalwart McCullochs of our day are bound to drift into one of those positions of which it is hard to say whether they involve renunciation, reinterpretation, or misunderstanding of the original message.

V

Our examples might suggest that analytically uncontrolled ideas play their role exclusively in the realm of those broad conceptions of the economic process as a whole that constitute the background from which analytic effort sets out and of which we never succeed in fully mastering more than segments. This is of course true to some extent – the bulk of our research work deals with particulars that give less scope to mere vision and are more strictly controlled by objective tests – but not wholly so. Take, for instance, the theory of saving which does appear in a wider context in the Keynesian system but might also, factually and theoretically, be treated by itself. From the time of Turgot and Smith – in fact from still earlier times – to the time of Keynes all the major propositions about its nature and effects have, by slow accretion, been assembled so that, in the light of the richer supply of facts we command today, there should be little room left for difference of opinion. It should be easy to draw up a summarizing (though perhaps not very exciting) analysis that the large majority of professional economists might accept as a matter of course. But there is, and always has been, eulogistic or vituperative preaching on the subject that, assisted by terminological tricks such as the confusion between saving and nonspending, has succeeded in producing a sham antagonism between the writers on the subject. Much emphasized differences in doctrine for which there is no factual or analytical basis always indicate, though in themselves they do not prove, the presence of ideological bias on one side or on both – which in this case hails from two different attitudes to the bourgeois scheme of life.

Another instance of sectional ideology of this kind is afforded by the attitude of many, if not most economists, toward anything in any way connected with monopoly (oligopoly) and cooperative price setting (collusion). This attitude has not changed since Aristotle and Molina although it has acquired a partially new meaning under the conditions of modern industry. Now as then, a majority of economists would subscribe to Molina's dictum: *monopolium est injustum et rei publicae injuriosum*. But it is not this value judgment which is relevant to my argument – one may dislike modern largest-scale business exactly as one may dislike many other features of modern civilization – but the analysis that leads up to it and the ideological influence that this analysis displays. Anyone who has read Marshall's *Principles*, still more anyone who has also read his *Industry and Trade*, should know that among the innumerable patterns that are covered by those terms there are many of which benefit and not injury to economic efficiency and the consumers' interest ought to be predicted. More modern analysis permits to show still more clearly that no sweeping or unqualified statement can be true for all of them; and that the mere facts of size, single-sellership, discrimination, and cooperative price setting are in themselves inadequate for asserting that the resulting performance is, in any relevant sense of the word, inferior to the one which could be expected under pure competition in conditions attainable under pure competition – in other words, that economic analysis offers no material in support of *indiscriminate* "trust busting" and that such material must be looked for in the particular circumstances of each individual case. Nevertheless, many economists support such *indiscriminate* "trust busting" and the interesting point is that enthusiastic sponsors of the private-enterprise system are particularly prominent among them. Theirs is the ideology of a capitalist economy that would fill its social functions admirably by virtue of the magic wand of pure competition were it not for the monster of monopoly or oligopoly that casts a shadow on an otherwise bright scene. No argument avails about the performance of largest-scale business, about the inevitability of its emergence, about the social costs involved in destroying existing structures, about the futility of the hallowed ideal of pure competition – or in fact ever elicits any response other than most obviously sincere indignation.

Even as thus extended, our examples, while illustrating well enough what ideology is, are quite inadequate to give us an idea of the range of its influence. The influence shows nowhere more strongly than in economic history which displays the traces of ideological premises so clearly, precisely because they are rarely formulated in so many words, hence rarely challenged – the subject of the role that is to be attributed in economic development to the initiative of governments, policies, and politics affords an excellent instance: groupwise, economic historians have systematically over- or understated the importance of this initiative in a manner that points unequivocally to prescientific convictions. Even statistical inference loses the objectivity that should in good logic characterize it whenever ideologically relevant issues are at stake.[7] And some of the sociological, psychological, anthropological, biological waters that wash our shores are so vitiated by ideological bias that, beholding the state of things in parts of those fields, the economist might sometimes derive solace from comparison. Had we time, we could everywhere observe the same phenomenon: that ideologies crystallize, that they become creeds which for the time being are impervious to argument; that they find defenders whose very souls go into the fight for them.

There is little comfort in postulating, as has been done sometimes, the existence of detached minds that are immune to ideological bias and *ex hypothesi* able to overcome it. Such minds may actually exist and it is in fact easy to see that certain social groups are further removed than are others from those ranges of social life in which ideologies acquire additional vigor in economic or political conflict. But though they may be relatively free from the ideologies of the practitioners, they develop not less distorting ideologies of their own. There is more comfort in the observation that no economic ideology lasts forever and that, with a likelihood that approximates certainty, we eventually grow out of each. This follows not only from the fact that social patterns change and that hence every economic ideology is bound to wither but also from the relation that ideology bears to that prescientific cognitive act which we have called vision. Since this act induces fact finding and analysis and since these tend to destroy whatever will not stand their tests, no economic ideology could survive indefinitely even in a stationary social world. As time wears on and these tests are being perfected, they do their work more quickly and more effectively. But this still leaves us with the result that some ideology will always be with us and so, I feel convinced, it will.

But this is no misfortune. It is pertinent to remember another aspect of the relation between ideology and vision. That prescientific cognitive act which is the source of our ideologies is also the prerequisite of our scientific work. No new departure in any science is possible without it. Through it we acquire new material for our scientific endeavors and something to formulate, to defend, to attack. Our stock of facts and tools grows and rejuvenates itself in the process. And so – though we proceed slowly because of our ideologies, we might not proceed at all without them.

Notes

1. This passage should be clear. But it may be as well to make its meaning more explicit. The misconduct in question consists, as stated, in "bending facts or logic in order to gain a point for either an ideal or an interest" *irrespective of whether a writer states his preference for the cause for which he argues or not.* Independently of this, it may be sound practice to require that everybody should explicitly state his "axiological premises" or the interest for which he means to argue whenever they are not obvious. But this is an additional requirement that should not be confused with others.
2. In particular, its acceptance is no prerequisite of the validity of the argument that is to follow and could have been set forth also in other ways. There are, however, some advantages in starting from a doctrine that is familiar to all and that needs only to be mentioned in order to call up, in the mind of the audience, certain essential notions in a minimum of time.
3. I should consider it an insult to the intelligence of my readers to emphasize that in particular this kind of relativism has nothing to do with Einsteinian relativity were it not a fact that there actually are instances of this confusion in the philosophical literature of our time. This has been pointed out to me by Professor Philipp Frank.
4. The contrary opinion that is sometimes met with is to be attributed to the simplified versions of the marginal-productivity theory that survive in text-books and do not take into account all the restrictions to which production functions are subject in real life, especially if they are production functions of going concerns for which a number of technological data are, for the time being, unalterably fixed – just as in elementary mechanics no account is taken of the complications that arise so soon as we drop the simplifying assumption that the masses of bodies are concentrated in a single point. But a marginal-productivity theory that does take account of restrictions which, even in pure competition, prevent factors from being paid according to their marginal productivities is still marginal-productivity theory.
5. Even there, so I have been reminded by Professor E. Hamilton, there is perhaps more to praise than there is to blame.
6. See especially S. N. H. Linguet, *La Théorie des Lois Civiles* (1767), and Marx's comments on him in Volume 1, pp. 77 *et seq.* of the *Theorien über den Mehrwert.*
7. I am not aware of any instances in which the rules of inference themselves have been ideologically distorted. All the more frequent are instances in which the rigor of tests is relaxed or tightened according to the ideological appeal of the proposition under discussion. Since acceptance or rejection of a given statistical result always involves some risk of being wrong, mere variation in willingness to incur such a risk will suffice, even apart from other reasons, to produce that well-known situation in which two statistical economists draw opposite inferences from the same figures.

TWELVE

Welfare Propositions of Economics and Interpersonal Comparisons of Utility

Nicholas Kaldor

Nicholas Kaldor (1908–1986) was born in Budapest and educated in Budapest, Berlin, and at the London School of Economics. In addition to an academic career, which was centered at Cambridge University, Kaldor served as an advisor to several governments and was instrumental in devising the value added tax (VAT). In this brief essay, originally published in 1939, he argues that the net benefit of a policy – the amount that "winners" would be willing to pay minus the amount that "losers" would need to be compensated – provides a measure of the capacity of an economy to satisfy preferences that does not require interpersonal comparisons or any judgment concerning the justice of different distributions. In a separate essay published in the same year, John Hicks defends the same idea, and assessment of alternatives in terms of net benefits is often called "the Kaldor-Hicks efficiency criterion."

In the December 1938 issue of the Economic Journal Professor Robbins returns to the question of the status of interpersonal comparisons of utility.[1] It is not the purpose of this note to question Professor Robbins' view regarding the scientific status of such comparisons; with this the present writer is in entire agreement. Its purpose is rather to examine the relevance of this whole question to what is commonly called "welfare economics." In previous discussions of this problem it has been rather too readily assumed, on both sides, that the scientific justification of such comparisons determines whether "economics as a science can say anything by way of prescription." The disputants have been concerned only with the status of the comparisons; they were – apparently – agreed that the status of prescriptions necessarily depends on the status of the comparisons.

This is clearly Mr. Harrod's view. He says:[2] "Consider the Repeal of the Corn Laws. This tended to reduce the value of a specific factor of production – land. It can no doubt be shown that the gain to the community as a whole

Economic Journal, vol. 49 (1939): 549–52. Reproduced by permission of Blackwell Publishing.

exceeded the loss to the landlords – *but only if individuals are treated in some sense as equal.* Otherwise how can the loss to some – and that there was a loss can hardly be denied – be compared with the general gain? If the incomparability of utility to different individuals is strictly pressed, not only are the prescriptions of the welfare school ruled out, but all prescriptions whatever. The economist as an adviser is completely stultified, and unless his speculations be regarded as of paramount aesthetic value, he had better be suppressed completely." This view is endorsed by Professor Robbins:[3] "All that I proposed to do was to make clear that the statement that social wealth was increased [by free trade] itself involved an arbitrary element – that the proposition should run, *if* equal capacity for satisfaction on the part of the economic subjects be assumed, *then* social wealth can be said to be increased. Objective analysis of the effects of the repeal of duties only showed that consumers gained and landlords lost. That such an arbitrary element was involved was plain. It seemed no less plain, therefore, that, here as elsewhere, it should be explicitly recognised."

It can be demonstrated, however, that in the classical argument for free trade no such arbitrary element is involved at all. The effects of the repeal of the Corn Laws could be summarised as follows: (i) it results in a reduction in the price of corn, so that the *same* money income will now represent a higher real income; (ii) it leads to a shift in the distribution of income, so that some people's (*i.e.*, the landlord's) incomes (at any rate in money terms) will be lower than before, and other people's incomes (presumably those of other producers) will be higher. Since aggregate money income can be assumed to be unchanged, if the landlords' income is reduced, the income of other people must be correspondingly increased. It is only as a result of this consequential change in the distribution of income that there can be any loss of satisfactions to certain individuals, and hence any need to compare the gains of some with the losses of others. But it is always possible for the Government to ensure that the previous income-distribution should be maintained intact: by compensating the "landlords" for any loss of income and by providing the funds for such compensation by an extra tax on those whose incomes have been augmented. In this way, everybody is left as well off as before in his capacity as an income recipient; while everybody is better off than before in his capacity as a consumer. For there still remains the benefit of lower corn prices as a result of the repeal of the duty.

In all cases, therefore, where a certain policy leads to an increase in physical productivity, and thus of aggregate real income, the economist's case for the policy is quite unaffected by the question of the comparability of individual satisfactions; since in all such cases it is *possible* to make everybody better off than before, or at any rate to make some people better off without making anybody worse off. There is no need for the economist to prove – as indeed he never could prove – that as a result of the adoption of a certain measure nobody in the community is going to suffer. In order to establish his case, it is quite sufficient for him to show that even if all those who suffer as a result are fully compensated for their loss, the rest of the community will still be better off than before. Whether the landlords, in the free-trade case, should in fact be given compensation or not, is a political question on which the economist, *qua* economist, could hardly pronounce an opinion. The important fact is that, in the argument in favour of free trade, the fate of the landlords is wholly irrelevant: since the benefits of free trade are by no means destroyed even if the landlords are fully reimbursed for their losses.[4]

This argument lends justification to the procedure, adopted by Professor Pigou in *The Economics of Welfare*, of dividing "welfare economics" into two parts: the first relating to production, and the second to distribution. The first, and far the more important part, should include all those propositions for increasing social welfare which relate to the increase in aggregate production; all questions concerning the stimulation of employment, the equalisation of social net products, and the equalisation of prices with marginal costs, would fall under this heading. Here the economist is on sure ground; the scientific status of his prescriptions is unquestionable, provided that the basic postulate of economics, that each individual prefers more to less, a greater satisfaction to a lesser one, is granted. In the second part, concerning distribution, the economist should not be concerned with "prescriptions" at all, but with the relative advantages of different ways of carrying out certain political ends. For it is quite impossible to decide on economic grounds what particular pattern of income-distribution maximises social welfare. If the postulate of equal capacity for satisfaction is employed as a criterion, the conclusion inescapably follows that welfare is necessarily greatest when there is complete equality; yet one certainly cannot exclude the possibility of everybody being happier when there is some degree of inequality than under a régime of necessary and complete equality. (Here I am not thinking so much of differences in the capacity for satisfactions between different individuals, but of the satisfactions that are derived from the prospect of improving one's income by one's own efforts – a prospect which is necessarily excluded when a régime of complete equality prevails.) And short of complete equality, how can the economist decide precisely how much inequality is desirable – *i.e.*, how much secures the maximum total satisfaction? All that economics can, and should, do in this field, is to show, given the pattern of income-distribution desired, which is the most convenient way of bringing it about.

London School of Economics

Notes

1. "Interpersonal Comparisons of Utility: A Comment," Economic Journal, December 1938, pp. 635–691.
2. "Scope and Method of Economics," *ibid.*, September 1938, pp. 396–397. (Italics mine.)
3. *Loc. cit.*, p. 638.
4. This principle, as the reader will observe, simply amounts to saying that there is no interpersonal comparison of satisfactions involved in judging any policy designed to increase the sum total of wealth just because any such policy *could* be carried out in a way as to secure unanimous consent. An increase in the money value of the national income (given prices) is not, however, necessarily a sufficient indication of this condition being fulfilled: for individuals might, as a result of a certain political action, sustain losses of a non-pecuniary kind – *e.g.*, if workers derive satisfaction from their particular kind of work, and are obliged to change their employment, something more than their previous level of money income will be necessary to secure their previous level of enjoyment; and the same applies in cases where individuals feel that the carrying out of the policy involves an interference with their individual freedom. Only if the increase in total income is sufficient to compensate for such losses, and still leaves something over to the rest of the community, can it be said to be "justified" without resort to interpersonal comparisons.

THIRTEEN

The Philosophical Foundations of Mainstream Normative Economics

Daniel M. Hausman and Michael S. McPherson

Michael S. McPherson (1947–) received his Ph.D. from the University of Chicago and taught for many years at Williams College before becoming President of Macalester College and then of the Spencer Foundation. His academic work concerns the economics of higher education and issues at the boundaries of economics and ethics. He and Daniel Hausman founded the journal *Economics and Philosophy* in 1985 and edited it for its first ten years. They also coauthored *Economic Analysis, Moral Philosophy, and Public Policy*, from which this essay derives.

Let us begin with an old joke. Brezhnev and other members of the Soviet Central Committee are reviewing a May Day parade in Moscow. Thousands of infantry march by, followed by armored cars, the latest tanks, long range artillery, and progressively larger, sleeker, and more impressive missiles. At the end, a battered flatbed truck rumbles by carrying a half-dozen unathletic and bespectacled middle-aged men and women in dirty raincoats sitting around a card table. The crowd is restless and members of the Central Committee are scandalized. One is bold enough to ask Brezhnev what these nondescript civilians are doing in the midst such a magnificent military parade. Brezhnev replies, "Ah, those are our economists. You'd be amazed at the damage they can do."

Like most economist jokes, this one is unkind, but its unkindness should not be exaggerated. It refers to the damage economists *can* do, not to any inevitable harm that they cause. And there is no suggestion that their intentions are evil. Economics can unfortunately do great harm, but we think that it can do good, too. It is a sharp two-edged sword that needs to be mastered

This essay derives from a long collaboration, which led to the publication of *Economic Analysis and Moral Philosophy* in 1996 and a second edition, *Economic Analysis, Moral Philosophy, and Public Policy* in 2006. In between, Hausman published an essay, "The Philosophical Foundations of Normative Economics," in Ayogu and Ross 2005, which borrowed from the first edition and strongly influenced this essay.

and handled with care. Economics has done harm mainly because it has been misunderstood or misused by political and economic interests. There is little that philosophers or economists can do to combat powerful interests who are ready to exploit any theory – economic, political, even philosophical – to rationalize their ambition and greed. But philosophers and economists can clarify the interpretation of economic theory and thereby remove confusions and make its ideological misapplication at least a little bit more difficult. Such is the objective of this essay with respect to mainstream normative economics. Our hope is that an understanding of the contestable assumptions upon which normative economists rely can help prevent people from being buffaloed by confident proclamations of economic "wisdom."

1. A Notorious Example

Let us begin with a notorious example to illustrate the central features of normative economics and to show why an inquiry into its philosophical foundations is needed. In December 1991, Lawrence Summers, then the chief economist at the World Bank, sent a memorandum to colleagues containing the following remarks:

> Just between you and me, shouldn't the World Bank be encouraging more migration of the dirty industries to the LDC's [less developed countries]? I can think of three reasons:
>
> (1) The measurement of the costs of health-impairing pollution depends on the foregone earnings from increased morbidity and mortality. From this point of view a given amount of health-impairing pollution should be done in the country with the lowest cost, which will be the country with the lowest wages. . . .
>
> (2) . . . I've always thought that under-populated countries in Africa are vastly *under* polluted; their air quality is probably vastly inefficiently [high] compared to Los Angeles or Mexico City. Only the lamentable facts that so much pollution is generated by non-tradable industries (transport, electrical generation) and that the unit transport costs of solid waste are so high prevent world-welfare-enhancing trade in air pollution and waste.
>
> (3) The demand for a clean environment for aesthetic and health reasons is likely to have very high income-elasticity. The concern over an agent that causes a one-in-a-million change in the odds of prostate cancer is obviously going to be much higher in a country where people survive to get prostate cancer than in a country where under-5 mortality is 200 per thousand. . . . The problem with the arguments against all of these proposals for more pollution in LDCs (intrinsic rights to certain goods, moral reasons, social concerns, lack of adequate markets, etc.) could be turned around and used more or less effectively against every Bank proposal for liberalisation. (quoted in *The Economist*, February 8, 1992, p. 66)

Summers was not seriously proposing a World Bank program to export pollution to the LDCs. This memorandum is of interest instead because Summers baldly put into words uncomfortable implications that most economists would prefer not to draw.

Summers's memorandum makes claims about what the World Bank "should" be doing, and it describes some facts as "lamentable." Summers is clearly making evaluative claims, and his work would be excluded from economics by those who insist that economics must be free of any value judgments. Yet this memorandum obviously seems to be concerned with economics. One way to recognize this, while still insisting on the importance of distinguishing between factual and evaluative claims, is to maintain that there are two kinds of economics: "positive economics," which deals only with matters of fact, and "normative economics," which is concerned with the evaluation of economic states of affairs, processes, and institutions. Summers's memorandum is clearly an instance of normative economics.

It is useful to distinguish seven features of Summers's memorandum, which are typical of mainstream normative economics or "welfare economics." Each of these features represents a *choice*: Summers's way of thinking about economic states of affairs and policies is just one of *many* possible ways. Once one recognizes what distinguishes this way of thinking about outcomes and policies from other ways, one understands a great deal about normative economics.

1. Summers is concerned with *evaluating* economic states of affairs and with recommending how to improve them. His focus is on economic outcomes rather than processes.
2. Summers assumes that there is a single framework for economic evaluation, which he takes for granted. He relies on an unstated ethical foundation that he believes his readers share.
3. The memorandum considers how policies and states of affairs bear on *individuals*. No questions are asked about the significance of their effects on other things such as the environment or local cultures, except insofar as those in turn affect the welfare of individuals.
4. The memorandum evaluates economic states of affairs exclusively in terms of their consequences for individual *welfare*. Because of the prevalence of this feature, mainstream normative economics is typically called "welfare economics."
5. In measuring welfare, the memorandum implicitly accepts the way that markets evaluate things.
6. Although the memorandum focuses exclusively on welfare, it does not add up welfare gains and losses or compare the welfare gains and losses

of different people. Summers does not claim that trade in pollution would maximize total or average welfare.

7. In addition to focusing exclusively on the welfare implications of shifting pollution, the memorandum suggests that there is a qualitative difference between the "impeccable" "economic *logic* of dumping a load of toxic waste in the lowest-wage country" [our emphasis] and miscellaneous and unspecified ethical objections in terms of "intrinsic rights to certain goods, moral reasons, social concern, lack of adequate markets, etc." Summers implies that the welfare arguments are rigorous and worth taking seriously, while the miscellaneous objections can be disregarded.[1] Although welfare economists rarely deny that other moral considerations are relevant to evaluating policies and outcomes, they are often suspicious, impatient, or even contemptuous of other ethical concerns.

Some of these seven features of welfare economics are widely shared in the thought and culture of modern liberal democracies, whereas others are more distinctive to mainstream economics. None of these features of welfare economics, even these that are widely shared with liberal social theory more generally, is inevitable. Each involves a choice, and each feature could be questioned or changed. These choices are both methodological and ethical. Although welfare is obviously very important, so is freedom and so is justice. Normative economics might focus on them in addition to or instead of welfare. There are alternatives, and to choose among them requires ethical reflection.

Table 13.1 lists some of the alternatives to the choices economists have made and helps to make clear what is distinctive about the standards of evaluation that normative economics relies on.

2. Summers's Argument

Air and water pollution lessen the quality of life in many ways, yet most kinds of pollution have no market prices, because it is difficult to locate the sources of some pollution and expensive to strike a deal with all the polluters in order to improve your air or water. In addition, any deal you strike with polluters will affect your neighbor and vice versa – while walking to the corner, you've got to breathe the same air your neighbor breathes. Any effective deal will require cooperation among your neighbors.

Thus, some collective action is often needed in controlling pollution. One way that economists can help with the problems of controlling pollution is by *imputing* costs to it. The hope is to figure out what pollution costs *would*

Table 13.1. *The Moral Framework of Normative Economics*

1. What should economists appraise?
   a. √*Outcomes*
   b. Processes
2. What method(s) of appraisal should economists use
   a. √*Single method of appraisal*
   b. Multiple ethical perspectives, depending on problem
3. What matters about outcomes?
   a. √*Consequences for individuals*
   b. Consequences for groups or the environment
4. Which features of outcomes for individuals matter?
   a. √*Welfare*
   b. Freedom
   c. Rights
   d. Justice
5. What is welfare?
   a. √*The satisfaction of preferences*
   b. Some mental state, such as happiness
   c. "Objective" goods; e.g. achievements, personal relations, health, etc.
6. How does welfare (as preference satisfaction) bear on the evaluation of outcomes?
   a. √*Market evaluation and the Pareto concepts*[2]
   b. Add up preference satisfaction
7. What role do other ethical notions play?
   a. √*Independent: important, but not a concern of economics*
   b. Their importance is derivable from their consequences for welfare
   c. Must be integrated into the economic appraisal
   d. Of no importance

be, if there were markets where pollution could be bought and sold. For example, economists may attempt to impute pollution costs by examining housing prices in communities that are much the same, apart from their air quality. Economists have a number of ingenious techniques by which they can estimate how much people in developed countries would be willing to pay to lessen pollution in their environment and how much people in LDCs would have to be compensated in order to be willing to accept more pollution.

Summers argues in addition that these measurements do not result from people's ignorance or mistaken beliefs. In his view, the economic costs of the consequences of increased pollution are *in fact* much lower in LDCs than they are in developed countries. Rational and well-informed people in LDCs *should be* happy to sell pollution rights to people in developed countries for a price that the latter *should be* happy to pay. The willingness to accept more pollution in LDCs does not rest on mistakes about the consequences of doing so.

Suppose that environmental quality could be bought and sold in individual privately consumable units and consider whether rational and well-informed individuals, who live in a particular LDC, *L*, could strike deals to sell units of "environmental quality" to rational and well-informed individuals, who live in a developed country, *D*. If *L* is one of those "underpolluted" LDCs that Summers refers to, it has a great deal of inexpensive environmental quality, whereas in *D*, by contrast, environmental quality is costly and scarce. So unless the price of a unit of environmental quality is extremely high or extremely low, individuals in *both L* and *D* will want to trade.

So if individuals were all rational and well informed, and it were possible for individuals easily to buy, sell, and transport pollution or "environmental quality," there would be active trading between the developed and less developed nations of the world, and pollution would be pouring out of the developed nations and into the less developed nations. This happy outcome is not feasible, because units of environmental quality cannot be individually appropriated, bought, and sold, and it is hard to transport pollution. Summers laments these barriers to trade, and he thinks the World Bank can enhance world welfare by helping to move pollution to LDCs in return for some measure of compensation.

Merely shifting pollution to LDCs, without paying any compensation could not, of course, be *mutually* beneficial, because the LDCs would be harmed. But it would still result in what economists call a "net benefit," because the developed countries *could* compensate the LDCs and still allegedly be better off. We shall discuss the notion of a net benefit and the justification for favoring policies that provide net benefits later in Section 7.

Why should Summers conclude that it is "lamentable" that "pollution is generated by non-tradable industries?" How does Summers reach the conclusion that "the World Bank [should] be encouraging *more* migration of the dirty industries to the LDC's?" How do normative economists get from claims about how rational and well-informed individuals *would choose* to claims about *welfare* and from claims about welfare to claims about what the World Bank *ought to do*? What is the logic of Summers's argument?

Here is one way to spell it out:

1. For some amount of compensation *C* between the least agents in LDC will accept and the most agents in rich countries will offer, all rational individuals, whether in developed countries or in LDCs, would prefer to transfer pollution from a developed country to a LDC. (premise)

2. Whatever well-informed and rational individuals prefer makes them better off. (premise)
3. So exporting pollution to LDCs from developed countries and paying compensation makes everyone better off. (from 1 and 2)
4. One should adopt policies that make people better off. (premise)
5. One should adopt policies that shift pollution to LDCs and pay compensation. (from 3 and 4)

If one assumes that the jobs and revenues provided by dirty industries are adequate compensation, then this reconstruction may capture Summers's intentions.

The tone of Summers's memorandum suggests that the three numbered paragraphs make a "scientific" case, whereas the last paragraph mentions wishy-washy moral objections. But the moral content does not wait for the last paragraph to make its appearance. As this reconstruction shows, the three numbered paragraphs are part of a *moral* argument. One of its moral premises (premise 2) is particularly important to the link between market evaluation and welfare. By identifying welfare and preference satisfaction and then relying on the connection that positive economics establishes between preferences and market prices, Summers can link premises about costs and demands to conclusions about what outcomes will enhance welfare.

The uproar caused by this memo suggests that most people are not willing to accept its conclusion. Why not? Why shouldn't the World Bank encourage migration of dirty industries? Here are five objections:

1. Encouraging dirty industries to migrate to LDCs might lead to more total pollution. Developed countries have stronger incentives, greater administrative capacity, and more resources to enforce pollution controls than do LDCs. This is an important objection, but it does not challenge Summers's framework, and we shall say nothing more about it.
2. Even if people in both developed economies and LDCs would prefer to shift pollution to LDCs in exchange for appropriate compensation, the exchange may be *unfair*. Developed countries are exploiting the poverty of LDCs – which, in addition, they are often responsible for. It may not be right to make people better off if doing so involves injustice. Justice matters, too.
3. Summers's analysis compares only one possible alternative to the status quo: that of shifting pollution to LDCs. But there may be other policies that would be better still. Notice in particular that Summers's case depends on the huge income disparities between rich and poor

countries: without those disparities, why would people in Nigeria pay less to avoid pollution than people in the United States? Should this status quo income disparity be simply taken as given? Transferring wealth from rich to poor countries might enhance welfare more effectively than transferring pollution.

4. Satisfying preferences does not automatically increase welfare. People may prefer things that are bad for them. Voluntary exchange is not always mutually advantageous.
5. Premise 1, that all rational and well-informed agents would prefer to make the exchange is also controversial. This premise is itself the conclusion of an argument from the fact that the (economic) costs of pollution are lower in LDCs than in developed countries. But do the economic costs and benefits capture what is morally relevant? Do *rational and well-informed individuals have to accept the market's evaluation of the consequences of the pollution*? Isn't premise 1 a controversial moral premise, too? Given the current unequal distribution of wealth, preventing or curing a crippling injury or a case of AIDS confers much greater economic benefits in rich countries than in poor ones. But the moral significance of crippling injuries or of AIDS should not depend on whether the victim lives in a wealthy country or on the victim's own current or prospective income or wealth. One can thus reasonably raise moral objections to regarding economic costs and benefits as a guide to what *ought* to be done. Costs and prices have a contestable moral significance built into them.

In fact, economists do not typically identify the value of a human life with the loss of expected earnings or with the differing amounts different people would pay to prevent a death, and in a serious argument for a World Bank initiative, Summers would probably not have done so. But why not? If economic costs and benefits are a good guide to what is harmful and beneficial, then they should be a good guide to the allocation of risks of death and injury; and if they are not a good guide to the allocation of risks of death and injury, why should one believe that they provide an acceptable way to measure benefits or harms? Summers reduces the question of whether LDCs are "underpolluted" to the question of whether the welfare consequences of shifting more pollution to the LDCs would be favorable. "Welfare" for Summers, as for most economists, is preference satisfaction. The "cost" of the consequences of pollution is thus the amount by which people's preferences are less well satisfied. And Summers's measure of preference satisfaction is willingness-to-pay.[3]

Although more provocative and transparent than most normative economics, Summers's memorandum exemplifies common features of mainstream economic evaluation. Normative economists typically attempt to offer policy advice while setting aside considerations such as "intrinsic rights to certain goods, moral reasons, social concerns." They focus exclusively on welfare, which they associate with preference and willingness to pay. So normative economics is welfare economics. Normative economists also typically make inferences concerning welfare on the basis of data concerning willingness to pay, and these inferences are inevitably biased toward the preferences of those who are rich.[4]

Let us then back up and spend some time with the philosophical foundations. Why do normative economists focus exclusively on welfare, and why are they committed to this theory of welfare?

3. Individualism, Rationality, and Self-Interest

At the core of both positive economics and welfare economics lie controversial commitments to *individualism* and to a particular view of *human nature*. In particular, one should distinguish three varieties of individualism: ontological, explanatory, and ethical individualism and two views of human nature: human nature as rational and human nature as self-interested.

In its simplest formulation, ontological individualism maintains that only mental states and physical objects, including human beings, are real. Cultures, social institutions, and so forth are not real. They must be understood instead as reifications of features of the physical environment or of the physical and mental states of people. Ontological individualism is untenable and difficult to formulate sensibly. We mention it only to distinguish it sharply from explanatory and ethical individualism.

Explanatory individualism (or what is often called "methodological individualism") can be interpreted in many ways. Sometimes, it is interpreted as the view that explanations of social phenomena that refer to social entities are at best provisional, if not downright objectionable. The form of explanatory individualism to which economists subscribe is less restrictive. Economists have no qualms about explanations that cite facts about prices, incomes, laws, or contracts, and all of these are, of course, social entities. The explanatory individualism that economists typically assume maintains that the fundamental explanatory principles or laws (apart from the laws of the natural sciences) should concern the preferences, beliefs, and choices of individual human beings. For example, an explanation of the effects of government fiscal policy that cites a particular value of the multiplier is acceptable only because economists believe that the value of the multiplier can – at least in principle – be explained in terms of individual preferences, beliefs, and choices given specific initial conditions. The explanatory individualism to which most economists subscribe concedes that social entities and facts have causal consequences, but it insists that those consequences are mediated by the beliefs, preferences, and choices of individuals. We are not sure whether this version of explanatory individualism is ultimately defensible.

Before turning in the next section to ethical individualism, let us explore how explanatory individualism interacts with the two theses concerning rationality to determine the broad outlines of mainstream economics. The first of the two theses concerning human nature is that human beings are *rational*. The core idea is that explanations of individual choices also often *justify* those choices. The factors that *cause* choices also function as *reasons* for choices. People act for reasons and it is typically possible to justify their actions in terms of their beliefs and preferences. If one conjoins this basic view of human nature with explanatory individualism, one arrives at the view that the central explanatory principles of economics should be principles of rational individual choice.

Accordingly, one finds that a theory of rationality lies at the heart of both positive and normative mainstream economics. Although many economists identify rationality and material self-interest, the official theory of rationality denies that any particular objective, such as self-interest, is any more or less rational than any other objective. The official theory of rationality is *formal*. Rationality lies in the *structure* or *form* of choice and preferences, not in the content of what is preferred or chosen.

The theory of rationality embedded in mainstream economics states that individuals choose (or act) rationally if their actions are determined by their preferences, and their preferences are themselves rational. In modeling beliefs as subjective probabilities, economists also accept an implicit theory of rational belief, which we shall not discuss here. Preferences are rational if they are complete and transitive. An agent's preferences are complete if they rank all the alternatives the agent faces. For any two options x and y, either the agent prefers x to y or y to x, or the agent is indifferent between x and y. An agent's preferences are transitive if the agent prefers x to z whenever the agent prefers x to y and y to z (and similarly for indifference). There are further technical issues, but the basic idea is that when an agent has rational preferences, then – regardless of the content of the preferences – the agent has a consistent preference ranking of all the alternatives among which the agent can choose. It is as if the objects of choice could be written down in a list, with the more preferred alternatives in higher rows and alternatives among which the agent is indifferent sharing the same rows. Numbers can be assigned to rows, and those numbers – which are only indices representing places in the preference ranking – are what economists call "ordinal utilities." A utility function is just an assignment of numbers to alternatives in a way that indicates preference. "Maximizing utility" is simply doing what one most prefers. Utility is not itself an object of preference. It is not something sought or traded off against other things, because it is not a thing at all.

The theory of rationality is a normative theory, although *not* by itself a moral theory. One's preferences can be as rational in the pursuit of evil as in the pursuit of good. If one fails to choose what one prefers, then one is foolish, not necessarily morally culpable. As a normative theory, the theory of rationality says how people *should* behave, not what people actually do. Behavior that conflicts with the theory may thus show only that people fail to act rationally, rather than revealing any mistake in the theory. But if people's choices and preferences were not approximately in accord with the standard theory of rationality, then that theory would have little use except as a basis for criticism; and those who take human nature to be fundamentally rational would grow suspicious of the theory. Mainstream economists in fact take the further step of assuming that people are in fact rational, at least to some reasonable degree of approximation.

The standard formal theory of rationality does not go very far in fleshing out explanatory individualism and a view of human nature as rational. Without any general claims about the *content* of people's preferences, very little can be predicted about how they will choose, and, in the wake of their choices, little can be said except that they chose as they preferred. Positive economics becomes contentful only when economists offer generalizations concerning *what* people prefer. The most important of these generalizations is that people are materially self-interested, that they prefer more commodities to fewer, more wealth to less wealth. This generalization is so important, that one might reasonably think of it as a second general principle of human nature to which most mainstream economists are committed.

In speaking of rationality and self-interest as general principles of human nature to which economists subscribe, we do not mean to suggest that economists regard these principles as exceptionless general laws. Just as one can in various ways qualify the claim that individuals are rational, so one can hedge the claim that they are self-interested. Economists may, for example, take self-interest to be a reasonable approximation rather than the literal truth. Economists can avoid dealing with the conflicts between self-interest and concern for one's family by "cheating" on explanatory individualism and treating agents as households rather than as individuals. And so forth.

With the addition of self-interest, the fundamental theory now has significant content. Add diminishing marginal utility (or diminishing marginal rates of substitution), the assumption that people are well informed, and subsidiary assumptions concerning, for example, the divisibility of commodities, and economists can use their fundamental theory to explain market phenomena such as the law of demand.

4. Moral Foundations: Ethical Individualism and Welfare

To put forward a theory of normative economics requires that economists say something about ethics. Here again, economists are committed to a form of individualism. *Ethical individualism* is the view that social entities are of no *intrinsic* moral importance. There is moral reason to protect a culture, a religion, a state, a tribe, or a corporation if and only if doing so is required by moral concern for individual human beings. Ethical individualism leaves open the possibility that nonhuman animals, or perhaps even plants, have intrinsic moral worth. It denies specifically that there is anything morally significant about the interests of social entities, unless their protection can be linked to concerns about individuals.

These days, ethical individualism is increasingly controversial, as many of those who defend multiculturalism in the United States and who oppose globalization across the world argue for the importance of protecting distinctive local cultures. But a recognition of the enormous value of local cultures and of the enormous harm that results from their disruption is not inconsistent with ethical individualism. Ethical individualists should value local cultures very highly when they benefit their members and do not harm outsiders, and the loss of cultural *variation* in the long run may be as harmful to those who belong to hegemonic cultures as to those who belong to endangered cultures. Treating the moral value of cultures, languages, or other social practices as instrumental rather than intrinsic is fully consistent with valuing them *extremely* highly, but it does mean that the ethical individualist has no moral regrets about the death of social practices that do not promote the rights, freedoms, and interests of individuals.

Most Western ethical theories endorse some version of ethical individualism. Utilitarianism adopts a particularly simple variant. According to the utilitarian, only the welfare of sentient beings matters morally. So social policies, processes, practices, and institutions should be appraised by their consequences for individual well-being. Utilitarians can nevertheless find room to value justice, equality, and individual rights and liberties, as all these things contribute to individual well-being. Mainstream welfare economics, which was in fact influenced by utilitarianism, at first glance appears to follow utilitarianism in reducing ethical individualism – which is a plausible and humane doctrine – to the more dubious view that only individual *welfare* is of intrinsic moral importance.

It is not very informative to say that individual *welfare* is the sole thing of intrinsic moral importance until one has spelled out what welfare is, and without some means of tracing the welfare consequences of policies and of measuring welfare, this view does not help to evaluate policies. Bentham took utility to be that property of objects that causes sensations of pleasure in us (Broome 1991). Mill took well-being to be "happiness," but it is far from clear what he took happiness to be (1863, chapter 3). Economists have not been eager to wade into these murky philosophical waters,[5] yet without some notion of what welfare might be and some way of measuring welfare, they would have no way to evaluate policies or to offer guidance to policy makers.

Economists have opted for the view that welfare can be measured by the extent to which preferences are satisfied. Although there is surely some connection between welfare and preference satisfaction, it is unjustifiable to identify them. If welfare were the satisfaction of preference, then it would be not only unusual to prefer to sacrifice one's own welfare to some other end; it would be logically impossible! If welfare were the satisfaction of preference, then smoking would benefit those who prefer to smoke even if their preferences depended on their ignorance of the consequences of smoking. If welfare were the satisfaction of preference, then we would be better off if, as we prefer, there are no nuclear wars in the twenty-sixth century, even though we will by then – alas – have been dead for centuries.

Why then do mainstream economists nevertheless identify welfare and preference satisfaction? There are many explanations. One mistaken way to link welfare and preference is to equivocate on the word, "utility," which is both the name that positive economists give to an index of the extent to which preferences are satisfied and the name that the utilitarians gave to that which morality aims to maximize. If one erroneously takes the word to refer to the same thing in both contexts, then one will conclude that welfare is the satisfaction of preferences. It is also easy to equivocate on the word, "satisfaction." A person's preference is satisfied if things are as the person prefers them to be, regardless of how well satisfied the person *feels*. Indeed, many preferences may be satisfied without the person even knowing. Yet it is easy to slide from the view that welfare is a mental state such as a feeling of satisfaction to the view that welfare is the satisfaction of preference. A third mistaken route to the position that welfare is the satisfaction of preferences rests on the confusion of this view of welfare with the condemnation of paternalism. If whatever people prefer is automatically better for them, then the question of whether it is justifiable to coerce people for their own good can never even arise. But there are better ways to object to paternalism than to maintain falsely that people never prefer harmful alternatives. In his famous critique of paternalism in *On Liberty* (1859), John Stuart Mill criticizes the view that people should be coerced when they make choices that frustrate their own ends. He does not argue that whatever people choose is automatically good for them.

There are also more respectable routes to the identification of welfare and the satisfaction of preference. Given the two basic theses concerning human nature that mainstream economists accept – that individuals are rational and that they are self-interested – people will prefer x to y if and only if they believe that they will be better off with x than with y. If one supposes in addition – as positive economists typically do – that people's beliefs are generally correct, then people will prefer x to y if and only if they are in fact better off with x than with y. Regardless of what welfare *is*, people's preferences will then be a good guide to what makes them better off.

5. Repudiating Interpersonal Comparisons

Mainstream normative economics is distinctive, because it focuses almost exclusively on welfare, because it measures welfare by preference satisfaction, and because it denies that it is possible to compare welfare levels or differences across people. This last feature distinguishes welfare economics sharply from utilitarianism, which judges policies by their consequences for total or average utility.

If one individual prefers x to y and another has the opposite preferences, then the first individual's preferences can be represented by a utility function that assigns a higher number to x than to y, and the second's preferences can be represented by a utility function that assigns a smaller number to x than to y. But if these are ordinal utility functions that merely represent preference ranking, then the magnitudes of the numbers are otherwise completely arbitrary. Any attempt to add up the utility indices to determine whether preferences are better satisfied with x or with y shows a misunderstanding of what the indices mean. The comparison of the utility sums depends on an arbitrary assignment of utility indices and says nothing at all about which alternative satisfies preferences better. If there is nothing more to say about preferences than an ordinal utility function says, then utility differences could not be compared even for a single individual, and utilities could never be added or subtracted or compared across individuals.

If the preferences of individuals satisfy the stronger axioms of expected utility theory, then it is possible to represent them with utility functions in which utility sums and differences are not arbitrary.[6] *If* one had, in addition, some way to compare the utilities of different individuals, then the way would be cleared to formulate a utilitarian welfare economics. But what sense can one make of comparisons of the extent to which the preferences of different individuals are satisfied? Economists have stressed the problems of getting evidence concerning interpersonal comparisons of preference satisfaction, while we would stress the problems of even making *sense* of such comparisons (see Hausman 1995). In any case, such comparisons are highly problematic, and positive economics makes no use of them. So it is easy to see why normative economists have been hesitant to make interpersonal utility comparisons.

Denying the possibility of making interpersonal welfare comparisons largely determines the character of mainstream (normative) economics. It undercuts any hope of developing a general ethical theory, such as utilitarianism, that can encompass other dimensions of moral appraisal, such as freedom, rights, equality, and justice, because all of these are concerned with the differing weights of claims of different persons, which cannot be addressed within a welfarist framework unless one can make interpersonal comparisons. Few economists have the temerity to follow Bentham and to condemn all other moral considerations as "rhetorical nonsense – nonsense upon stilts." So some strategy is inevitable whereby economists appraise policies, outcomes, and institutions "other things being equal" or "along just one among several moral dimensions."

The idea that there is, thus, a specifically economic dimension of evaluation determines the character of mainstream normative economics. It is this idea that makes it possible to envision a normative economic *theory*, as opposed to a set of normatively motivated inquiries into consequences and properties of economic policies and institutions. Welfare economics depends not only on a specific view of welfare but also on the view that inquiries into welfare can be separated from inquiries into freedom, rights, equality, and justice. In one way, this separation limits economists. They can only appraise policies along one dimension or in one regard. But it also frees them from having to be concerned with anything but welfare. Having passed the buck with respect to everything except preference satisfaction, economists have only too often felt themselves free to ignore all other moral questions and to exaggerate the significance of their own partial mode of evaluation. But even without this exaggeration, the notion that there is a separate dimension of economic evaluation is, as we shall see, questionable.

6. Pareto Improvements, Pareto Efficiency, and Welfare Theorems

Appraising policies, outcomes, and institutions in terms of welfare without the possibility of making interpersonal comparisons is like running a race with your feet tied together. Just as it is possible awkwardly to hop 100 or 200 meters, so one can judge that *x* is better than *y* if somebody prefers *x* to *y* and nobody prefers *y* to *x*. In this case economists say that *x* is Pareto superior to *y* or that *x* is a Pareto improvement over *y*. The judgment that Pareto improvements are, other things being equal, moral improvements requires an additional weak moral judgment (which is arguably implicit in ethical individualism) that it is a good thing to make people better off. But even granting that additional judgment, the unanimity in preference required by the Pareto standard is seldom available. In serious policy debates none of the alternatives are Pareto superior to any of the others.

In addition to the notion of Pareto superiority, economists also define the notion of a Pareto optimum or of a Pareto efficient state of affairs. *x* is Pareto optimal or Pareto efficient if and only if there is no alternative that is Pareto superior to *x*. Notice *x*'s being Pareto optimal does not imply that *x* is Pareto superior to the alternatives, even to nonoptimal alternatives. Suppose, for purposes of illustration, that there are two people *A* and *B* and ten units of bread to distribute between them and that both *A* and *B* prefer more units to fewer, regardless of how many units they have. Then *any* distribution of bread to individuals that does not waste any bread is a Pareto optimum, but none of these Pareto efficient states of affairs is Pareto superior to any of the others. Furthermore, consider an inefficient state of affairs, such as one in which both *A* and *B* get four units and two units rot. Distributions whereby both get five units or one gets four and the other gets six are Pareto superior to the distribution where both get four. But a Pareto efficient state of affairs in which one gets nine units and the other gets one is not Pareto superior to the inefficient state of affairs in which both get four.

To say of a state of affairs that it is Pareto efficient is thus to express faint praise. The only thing praiseworthy about a Pareto efficient state of affairs is that it is not subject to one sort of criticism: It does not pass up any opportunities to satisfy some people's preferences better without sacrificing the preference satisfaction of somebody else. The fact that a Pareto efficient state of affairs is not faulty in this way may count for very little. In the simple example given here, if individuals will starve if they do not have at least two units of bread, the inefficient state of affairs where both get four units is much better than the efficient state of affairs where one gets nine units and the other gets only one.

Let us emphasize that we are here questioning the moral importance of the theoretical notion of Pareto optimality or Pareto efficiency. We are not questioning the moral importance of efficiency. Inefficiencies mean that fewer needs can be met. In harsh circumstances, inefficiencies mean more suffering. But there is a huge difference between a recognition of the importance of efficiency and an infatuation with Pareto optimality.

Mainstream economists have linked Pareto efficiency to competitive market equilibrium in two general welfare theorems. The first maintains that perfect competition guarantees Pareto efficiency. Insofar as Pareto efficiency is a good thing, so is perfect competition. The second theorem says that any distribution of income can be achieved as a perfectly competitive market outcome given the "right" initial distribution of resources. So, rather than regarding society as facing a trade-off between the efficiency provided by the market and various moral concerns about equity, one can have both. Concerns about equity can be met by fixing the initial distribution, and the market can then be relied on to bring about an efficient outcome.

Readers should not be overly impressed with either of these theorems. The fact that perfect competition guarantees Pareto efficiency is of little moral importance, first because Pareto efficiency is not a big deal, morally speaking. Inefficiency means that certain kinds of attractive improvements could in principle be made. Efficiency means that those sorts of improvements are not to be had, but nothing more. A second reason why people ought not to be overly impressed with the first welfare theorem is that perfect competition is impossible. A third reason is that, as Lipsey and Lancaster established (1956), the efficiency of perfect competition does not justify attempts to eliminate particular impediments to actual competition. (To grasp the intuition, consider a rough example: the distortions present in an economy in which half the output was in monopolized industries might be lessened by monopolizing the other half, even though making all industries competitive would be still better.) Unless one were to achieve perfect competition, which is impossible, eliminating market imperfections may well *diminish* rather than improve efficiency. There is no way to know, *a priori*.

Similar considerations undermine the significance of the second welfare theorem. Even if massive "initial" redistribution were politically feasible, perfect competition remains impossible, and eliminating market imperfections does not necessarily improve efficiency. We do not mean to underestimate either the actual magnitude of inefficiencies in the world or the moral importance of avoiding these inefficiencies, but the two welfare theorems are of no help in identifying or eliminating real inefficiencies. Who cares whether a perfectly competitive economy, which is not possible, is Pareto efficient and whether such an impossible economy could, as a consequence of an infeasible initial distribution, also result in an equitable distribution of income?

Mainstream economists care. Whether opposed to government intervention in the economy or in favor of it, mainstream economists typically treat perfectly competitive equilibrium as a benchmark and a moral ideal. Although some economists are opposed to government intervention in the economy for nonwelfarist reasons such as concerns about individual freedoms or rights, those who seek to limit the role of government usually maintain that freeing the economy from government meddling best approximates perfect competition with its Pareto efficiency. Those who, in contrast, favor government intervention in the economy do so because they believe that government can address some of the failures of actual markets, where these failures are identified against the ideal standard of perfect competition. Defenders of government interference with the market are just as impressed with the claims of perfect competition as are opponents, and they are both wrong to be so impressed.

7. Potential Pareto Improvements and Cost-Benefit Analysis

As unjustifiable as this fascination with perfect competition may be, it is easy to understand, because concern with competitive equilibrium is so central to positive economic theory, and because normative economists would have so little to say if they confined themselves to endorsing specific Pareto improvements. The only other path has been to find some way of comparing policies when none is Pareto superior to the others. Kaldor (1939) and Hicks (1939) had the following thought: Consider two economic outcomes or states of affairs *X* and *Y*. There are many different moral comparisons people might make of them. One morally significant difference between *X* and *Y* may be distributional, as in the case of the ten units of bread. Another way that economic states of affairs may differ is in the quantity of economic benefits to be distributed – that is, in their capacity to satisfy preferences. Suppose that in *X*, the status quo, four units of bread go to *A* and six units go to *B*. A new policy is considered that would increase bread supply and result in *A* getting seven units of bread and *B* getting five. Call the alternative *Y*. *Y* is not a Pareto improvement over *X*, because *B* gets fewer units of bread, but there is, Kaldor and Hicks argue, an unambiguous increase in economic benefits and economic efficiency. With the new policy, the capacity of the economy to satisfy preferences has increased. The "pie" has grown larger. That increase does not show up as a Pareto improvement, because there is also change in distribution to *B*'s disadvantage. Owing to the way the pie has been cut, *B*'s portion diminishes. In Kaldor and Hicks's view, economists are in no position to pass moral judgments on economic distribution, but they do not have to. The increase in efficiency, the purely economic benefit, is independent of distribution. Economists should be concerned to enlarge the pie, and they should leave its division to politicians and moralists. There is a separate dimension of purely economic evaluation.

According to Kaldor and Hicks, *X* has a greater capacity to satisfy preferences than *Y* if and only if *X* is a "potential Pareto improvement" over *Y*. *X* is a potential Pareto improvement over *Y* if there is some (not necessarily feasible) way of redistributing the goods available in *X* that makes *X* an actual Pareto improvement over *Y*. So, in the simple bread example, the distribution of seven units to *A* and five to *B* is a potential Pareto improvement over the distribution of four units to *A* and six to *B*, because it would be possible to redistribute the twelve units so as to achieve an actual Pareto improvement. (For example, both *A* and *B* could receive six units of bread.)

One also can describe a potential Pareto improvement in terms of the possibility of "compensation": if *X* is a potential Pareto improvement over *Y*, then it is possible in some sense for the winners in a change from *Y* to *X* to compensate the losers. Whether the winners could compensate the losers is then operationalized in terms of willingness to pay. If the amount that winners would be willing to pay to bring about a policy is larger than the amount that losers would need to be compensated to accept the policy, then the policy is a potential Pareto improvement over the status quo, and the policy purportedly brings about a more efficient state of affairs in which there is a "net benefit" – a greater capacity to satisfy preferences. All things considered, the policy might be a bad thing, because of its distributional consequences. But the distributional questions are not questions with which economists have any particular expertise. Furthermore, if the problems are distributional, then so are the solutions. The judgment concerning economic efficiency stands.

There are many problems with this argument, and many of these problems are inherited by the practical implementation of this line of thought in contemporary cost-benefit analysis. The central problem is that the separation that Kaldor and Hicks envisioned between questions concerning efficiency and distribution, between the size of the pie and the way it is sliced is not in general to be had.

Kaldor and Hicks's analysis would work if the utility frontiers were like those depicted in Figure 13.1, which represents the case of the loaves of

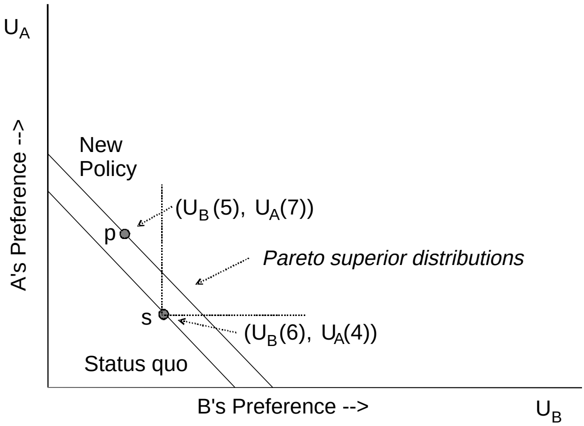

Figure 13.1. Pareto Superiority.

bread on the unrealistic assumption that *A* and *B*'s utilities are proportional to their bread consumption. The distribution resulting from the new policy, the point *p*, is not a Pareto improvement over the status quo distribution *s*, as B's utility is lower. But one can move along the frontier made possible by the new policy to a region of Pareto superior distributions.

This is a special case. There is no reason to rule out the situation depicted in Figure 13.2, which is borrowed from Samuelson (1950). The utilities of two representative individuals Rachel (R) and Peter (P) are measured along

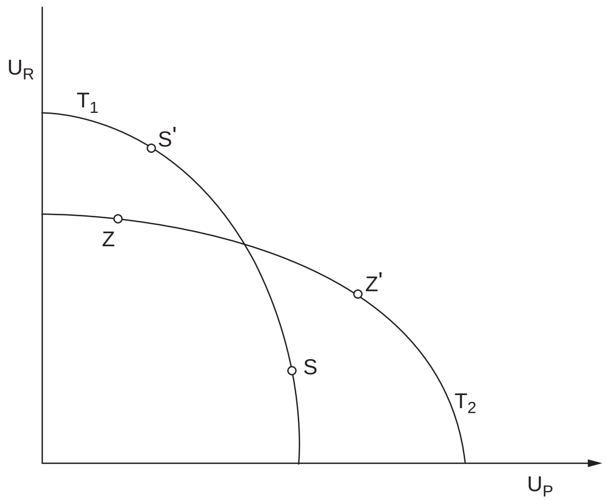

Figure 13.2. The Failure of Asymmetry.

the two axes, and the two curves represent the possible utility combinations depending on whether technology T_1 or technology T_2 is employed. Technology 1, which is the status quo (perhaps in 1920), involves a rail transport system with few goods or people carried on highways. Technology 2 involves an extensive road system like the one currently in use in the United States. The utility status quo in 1920 is, let us suppose, point *S*, and the result of changing to an automobile technology will, other things being equal, result in the utilities for Rachel and Peter represented by point *Z*. *Z* is not a Pareto improvement over *S*, because Peter is worse off, but *Z* is a potential Pareto improvement over *S*, because Rachel can pay compensation to Peter and the economy can move along the T_2 curve to *Z*. However, one cannot conclude, as Kaldor and Hicks hoped, that T_2 is more efficient than T_1, or that it involves a greater capacity to satisfy preferences, a larger real income, a bigger "pie," because *S* here is also a potential Pareto improvement over *Z*– one can move from *S* along the T_1 curve to *S*, which is an actual Pareto improvement over *Z*.

In just the same way, the efficiency of relocating polluting industries in LDCs that Summers points to depends in large part on the lopsided distribution of wealth. Indeed one can interpret "R" and "P" as standing for "rich" and "poor" and take T_2 to be the current technology (with polluting industries located in rich countries), *Z* the current level of preference satisfaction for representative members of rich and poor countries, and *S* the Pareto improvement that relocating the polluting industries will make possible. Although *S* is a Pareto improvement over *Z*, it cannot be compared to *Z*, which could be achieved by redistribution without shifting polluting industries. Only in the case in which one utility possibility curve is inside of the other can one make a "pure" efficiency comparison that does not take a particular distribution for granted. What makes *S* a Pareto improvement over *Z* is not a larger pie. *S* is a Pareto improvement over *Z* because of where *S* and *Z* sit on their respective frontiers. The greater "efficiency" of *S* depends on the distribution as well as the frontiers. Endorsing Pareto improvements is not neutral with respect to distributional questions.

In addition to the failure of the strategy of separating a specifically economic dimension of evaluation concerned exclusively with efficiency, let us briefly mention five other ethical problems with cost-benefit analysis. First, its appraisals are based on a comparison of "willingness-to-pay" rather than of welfare gains or losses of different people. Although willingness-to-pay obviously has something to do with welfare and preference, it also depends on expectations concerning what it is appropriate to purchase and for what price. Willingness to pay, like the amount of money one would require in order to consent to an unwanted change, obviously also depends on wealth. Because preferences in cost-benefit analysis are weighted with dollars, and the poor have fewer of these, their preferences count for less (Baker 1975).

Second, cost-benefit analysis, like other methods of evaluation employing the Pareto criteria, ignores questions of justice, even though, unlike the endorsement of Pareto improvements, it supports policies that make some people worse off. The compensation considered is only hypothetical. Some people win and some lose. Questions of fairness are obviously pressing in such circumstances. If each policy had different winners and losers so that in the long run everyone were a winner as often as he or she were a loser, the unfairness of individual policies taken separately might wash out. But the bias built into cost-benefit analysis against the preferences of the poor suggests that the unfairness will not wash out. Exactly those people whom policy makers should be most concerned to protect are those who are most likely to be harmed. Proponents have consequently explored ways of modifying cost-benefit analysis to compensate for possible injustices (Harberger 1978; Little 1957), but, in practice, wealth adjustments are seldom made.

A third objection to cost-benefit analysis is that social policy should not be based on the unreflective and unargued preferences that cost-benefit analysts infer from people's economic choices. Some preferences, such as preferences for communities free of urban sprawl, are hard to signal when one buys groceries, cars or even homes. Furthermore, people's preferences for public goods of all sorts respond to *arguments* and may be different after public debate than they were before. Substituting cost-benefit analysis for public deliberation means that people's preferences are never subjected to such challenges. Preferences that are based on mistaken beliefs wind up with the same influence on social policy as well-considered and well-informed preferences.

Finally, uncertainty coupled with the fact that preferences and willingness to pay typically depend on beliefs create serious problems. When people have mistaken beliefs about the constitution of the exhaust from the factory down the road, their willingness to pay to avoid breathing it will be an unreliable indicator of their true preferences, let alone the welfare consequences of the exhaust. People often do not know the consequences of alternatives and hence which alternative they would prefer if they did know the consequences. The problems of uncertainty are usually finessed by supposing that individuals possess subjective probability distributions over all the possible outcomes, but to suppose this involves extreme idealization; and there is little justification for respecting preferences based on largely fictitious probability distributions.

These problems do not imply that cost-benefit analysis is worthless or that it should be abandoned. But they do imply that it must be used with a great deal of circumspection. One also must abandon the hope of extracting a "purely economic" realm of evaluation, in which moral questions about distribution can be set aside. Information concerning willingness to pay can still help to decide what to do, but it is no more than one input into the messy business of policy making, rather than purportedly capturing a distribution-free notion of economic benefit.

It is still possible to make some reasonable guesses concerning the welfare consequences of alternative policies, and we believe that economists have an important role to play. But it is easy now to see both how much harm economists can do and the challenges that must be met in order for them to do good. In particular, economists need to surrender the view that they can focus on welfare alone and that preferences are always a reliable guide to welfare.

8. Conclusions

In making cautious use of the findings of a cost-benefit analysis, policy makers need to ask how dependent the net benefits are on the existing distribution of income and wealth and whether greater benefits might be obtained through redistribution. Recall Larry Summers's memorandum. It is plausible that redistributing income from rich nations to poor nations would increase overall well-being much more than redistributing pollution. And once one begins thinking of "overall" well-being, one has left behind cost-benefit analysis and the futile hope that "economic" questions can be sharply separated from distributional questions. The notion that welfare economists can offer a precise "economic" analysis to which vague moral concerns about justice or rights can be counterpoised is a pernicious fiction. The evaluation given by the market or simulated by welfare economists depends on a highly contestable theory of welfare and is no more solid or objective than other sorts of moral appraisals.[7] There are no short cuts to policy appraisal, and appraisals that ignore the full range of moral considerations bearing on economic policy are dangerous.

Notes

1. Otherwise, Summers's claim that these objections "could be turned around and used more or less effectively against every Bank proposal for liberalisation" should lead him to criticize the World Bank's proposals for liberalisation rather than to make an additional one!

2. These will be explained in Section 7 of this chapter.
3. Summers's other argument does not have this flaw. In some cases, a given exposure to a pollutant will in fact diminish the health and welfare of people in LDCs less than it will diminish the health and welfare of people in rich countries. If, to use Summers's own example, a pollutant increases the risk of prostate cancer, which is a disease mainly of elderly men, then the pollutant will not increase risk of suffering or death as much, if few men live long enough to contract the disease. Furthermore, for purely medical reasons, a given dose of a particular pollutant may have fewer negative health consequences if the total amount of pollution is small, than if it is large. Although these differential effects might not exist if there were not other inequalities between developing nations and LDCs, claiming that such pollution has a lower cost in LDCs does not involve valuing the lives of those who live in LDCs less. The weight of these arguments can be questioned, however, as the consequences of increased pollution may lie many years in the future, when the differences in longevity and levels of background pollution on which the differences in effect depend may have disappeared. Furthermore, the interaction between pollution effects and the generally worse health status of people in LDCs might render some of the effects of pollutants more rather than less serious. The numbers of people affected by pollution also must be considered. Adding up all these factors, there seems to be no justified presumption that transfers of pollution toward poor countries is morally desirable. It is thus questionable whether thoughtful people should or would be willing to transfer pollution from developed to developing countries. One further problem also should be mentioned. The idea of compensating a country is a cheat: to claim that everyone would be willing to transfer pollution illegitimately treats countries as if they were individuals. Even in its own terms the argument does not go through, because the compensation may fail to reach the individuals who are harmed by the pollution.
4. An extreme example of this is the drug eflornithine, which is a highly effective "miracle" cure for sleeping sickness. Until 1999, the drug was produced by a U.S. subsidiary of the Aventis company, but when eflornithine proved ineffective against cancer (its intended target), Aventis stopped making the drug and gave the production license to the World Health Organization (WHO). Only in early 2001, when stocks of the drug were almost exhausted, was the WHO able to find drug companies to manufacture it – and then only because the companies hoped to make profits from marketing eflornithine in developed countries as a cream that removes facial hair. Because the victims of sleeping sickness are so poor, the small amount they are able to pay for eflornithine grossly understates its social value as a cure for sleeping sickness.
5. Although recently there has been a flurry of interest in hedonist views. See Kahneman 2000, Kahneman and Krueger 2006, and Layard 2005.
6. More precisely, these "cardinal" utility functions are provably unique up to a positive affine transformation. This means that if one cardinally significant utility function U represents my preferences, then another cardinally significant utility function U will represent my preferences if and only if $U = aU + b$, where a is any positive real number and b is any real number.

7. Provided that one takes account of distributional presuppositions and consequences and recognizes that cost-benefit analysis is a source of data rather than answers, cost-benefit analysis can provide useful inputs into economic decision making. Of course, it is subject to abuse and misinterpretation, and the techniques employed to correct for distributional effects and to impute willingness to pay information from market data are certainly imperfect. But, provided that we don't forget that there are other things that matter besides welfare (let alone willingness to pay), what do we have that is less biased, more accurate, or that provides a better insight into what will serve people's material interests?

References

Ayogu, Melvin and Don Ross. 2005. *Development Dilemmas: The Methods and Political Ethics of Growth Policy*. London: Routledge.

Baker, C. 1975. "The Ideology of the Economic Analysis of Law." *Philosophy and Public Affairs* 5: 3–48.

Broome, John. 1991. "Utility." *Economics and Philosophy* 7: 1–12.

Harberger, Arnold. 1978. "On the Use of Distributional Weights in Social Cost-Benefit Analysis." *Journal of Political Economy* 86: s87–s120.

Hausman, Daniel. 1995. "The Impossibility of Interpersonal Utility Comparisons." *Mind* 104: 473–90.

Hausman, Daniel and Michael McPherson. 1996. *Economic Analysis and Moral Philosophy*. Cambridge: Cambridge University Press.

———. 2006. *Economic Analysis, Moral Philosophy and Public Policy*. Cambridge: Cambridge University Press.

Hicks, John. 1939. "The Foundations of Welfare Economics." *Economic Journal* 49: 696–712.

Kahneman, Daniel. 2000. "Experienced Utility and Objective Happiness: A Moment-Based Approach." pp. 673–692 of D. Kahneman and A. Tversky (Eds.) *Choices, Values and Frames*. New York: Cambridge University Press.

Kahneman, Daniel and Alan Krueger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20: 3–24.

Kaldor, Nicholas. 1939. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility." *Economic Journal* 49: 549–52.

Layard, Richard. 2005. *Happiness: Lessons from a New Science*. New York: Penguin.

Lipsey, Richard and Kelvin Lancaster. 1956. "The General Theory of the Second Best." *Review of Economic Studies* 24: 11–31.

Little, Ian. 1957. *A Critique of Welfare Economics*. 2nd ed. Oxford: Oxford University Press.

Mill, John Stuart. *On Liberty*. (1859). Ed.: Currin V. Shields. New York: Macmillan, 1985.

———. *Utilitarianism* (1863). Rpt. in Marshall Cohen, ed. *The Philosophy of John Stuart Mill*. New York: Modern Library, 1961, pp. 321–98.

Samuelson, Paul. 1950. "Evaluation of Real National Income." *Oxford Economic Papers New Series* 2: 1–29.

FOURTEEN

Why Is Cost-Benefit Analysis So Controversial?

Robert H. Frank*

Robert H. Frank (1945–) holds a B.S. in math from Georgia Institute of Technology and an M.A. in statistics and a Ph.D. in economics from University of California, Berkeley. Since serving as a Peace Corps volunteer in Nepal, he has been an economics professor at Cornell University. His "Economic Scene" column appears monthly in the *New York Times*. His work at the intersection of economics and ethics has focused on moral emotions and concerns about relative position.

Overview

The cost-benefit principle says we should take those actions, and only those actions, whose benefits exceed their costs. For many, this principle's commonsensical ring makes it hard to imagine how anyone could disagree. Yet critics of cost-benefit analysis are both numerous and outspoken. Many of them argue that cost-benefit analysis is unacceptable as a matter of principle. I begin by noting why many find this argument largely unpersuasive. I then examine several conventions adopted by cost-benefit analysts that do appear to yield misleading prescriptions. Finally, I consider the possibility that the cost-benefit principle may itself suggest why we might not always want to employ cost-benefit analysis as the explicit rationale for our actions.

The Incommensurability Problem

The cost-benefit principle says we should install a guardrail on a dangerous stretch of mountain road if the dollar cost of doing so is less than the implicit

* Henrietta Johnson Louis Professor of Management and Professor of Economics, Cornell University, Johnson Graduate School of Management. This paper was prepared for presentation at the University of Chicago Law School conference Cost-Benefit Analysis, September 17–18, 1999. I thank William Schulze for helpful discussions.

Journal of Legal Studies, vol. 29 (2000): 913–30.

dollar value of the injuries, deaths, and property damage thus prevented. Many critics respond that placing a dollar value on human life and suffering is morally illegitimate.[1]

The apparent implication is that we should install the guardrail no matter how much it costs or no matter how little it affects the risk of death and injury.

Given that we live in a world of scarcity, however, this position is difficult to defend. After all, money spent on a guardrail could be used to purchase other things we value, including things that enhance health and safety in other domains. Since we have only so much to spend, why should we install a guardrail if the same money spent on, say, better weather forecasting would prevent even more deaths and injuries?

More generally, critics object to the cost-benefit framework's use of a monetary metric to place the pros and cons of an action on a common footing. They complain, for example, that when a power plant pollutes the air, our gains from the cheap power thus obtained simply cannot be compared with the pristine view of the Grand Canyon we sacrifice.

Even the most ardent proponents of cost-benefit analysis concede that comparing disparate categories is extremely difficult in practice. But many critics insist that such comparisons cannot be made even in principle. In their view, the problem is not that we do not know how big a reduction in energy costs would be required to compensate for a given reduction in air quality. Rather, it is that the two categories are simply incommensurable.

This view has troubling implications. In the eyes of the cost-benefit analyst, any action – even one whose costs and benefits are hard to compare – becomes irresistibly attractive if its benefits are sufficiently large and its costs are sufficiently small. Indeed, few people would oppose a new technology that would reduce the cost of power by half if its only negative effect were to degrade our view of the Grand Canyon for just one 15-second interval each decade.[2] By the same token, no one would favor adoption of a technology that produced only a negligible reduction in the cost of power at the expense of a dark cloud that continuously shielded North America from the rays of the Sun. We live in a continuous world. If the first technology is clearly acceptable, and the second clearly unacceptable, some intermediate technology is neither better nor worse than the status quo. And we should count any technology that is better than that one as an improvement.

Scarcity is a simple fact of the human condition. To have more of one good thing, we must settle for less of another. Claiming that different values are incommensurable simply hinders clear thinking about difficult tradeoffs.

Notwithstanding their public pronouncements about incommensurability, even the fiercest critics of cost-benefit analysis cannot escape such tradeoffs. For example, they do not vacuum their houses several times a day, nor do they get their brakes checked every morning. The reason, presumably, is not that clean air and auto safety do not matter, but that they have more pressing uses of their time. Like the rest of us, they are forced to make the best accommodations they can between competing values.

General Reservations about Consequentialist Ethics

Many critics of cost-benefit analysis fault it for being rooted in utilitarianism or some closely related form of consequentialist ethical theory.[3] Consequentialist theories hold that the right course of action is the one that leads to the best consequences, where "consequences" under the utilitarian variant means "highest total utility." Critics often attack consequentialism by citing examples in which its purported conclusions clash with the reader's ethical intuitions. One popular example invokes the "utility monster," someone who transforms resources into utility far more efficiently than anyone else. Critics argue that since utilitarianism says the best outcome is to give all resources to the utility monster, and since we know this to be an absurd conclusion, we must reject the ethical theory upon which cost-benefit analysis rests.

Consequentialist moral philosophers have attempted to show that their theories, properly construed, do not imply the conclusions suggested by such examples.[4] But even if these disputes are never fully resolved, we may note that the theories favored by the rival camps reach remarkably similar decisions regarding a broad range of ethical questions. As a practical matter, then, the mere fact that cost-benefit analysis is closely identified with consequentialist ethical theories would not seem to imply that its prescriptions are systematically misleading.

Discounting the Future

As traditionally implemented, cost-benefit analysis attempts to put all relevant costs and benefits on a common temporal footing. A discount rate is chosen, which is then used to compute all relevant future costs and benefits in present-value terms. Most commonly, the discount rate used for present-value calculations is an interest rate taken from financial markets.

Though some critics complain about this practice, use of a market interest rate to discount future monetary costs and benefits commands broad approval. After all, if the annual interest rate on financial deposits is 7 percent, one can cover a $1,000 cost 10 years hence by depositing only $500 today.

There is less widespread agreement about using a market interest rate to discount future subjective utility. As Stanley Jevons argued, for example, "To secure a maximum benefit in life, all future pleasures or pains, should act upon us with the same force as if they were present, allowance being made for their uncertainty . . . But no human mind is constituted in this perfect way: a future feeling is always less influential than a present one."[5]

On this view, if failure to adopt more stringent air quality standards today means that respiratory illnesses will be more common a generation from now, those illnesses should receive roughly the same weight as if they were to occur today. Having been born later should not mean that one's enjoyment and suffering receive less weight in important policy decisions. Of course, a complete cost-benefit calculation would also want to make allowance for possible improvements in medical technology that would make the consequences of a given illness less severe in the future.

Whatever the ultimate merits of this position, it does not argue against the use of cost-benefit analysis as a matter of principle. If analysts agree that future experiences should receive roughly the same weight as current ones, the costs and benefits associated with any policy change can simply be calculated on that basis.

Distributional Issues

Distributional issues have long been a favorite target of critics of cost-benefit analysis. Their objection, in a nutshell, is that because willingness to pay is based on income, cost-benefit analysis assigns unjustifiably large decision weight to high-income persons. Implicit in this objection is the view that everyone's preferences regarding policy decisions should receive the same weight, irrespective of income.

Critics presumably have the interests of the poor in mind when they press this objection. Yet it is not clear that the poor themselves would want policy decisions to be made on some basis other than willingness to pay. Consider, for example, a community consisting of three voters – one rich, the other two poor. Up for decision is a proposal to switch the local public radio station from an all-music format to an all-talk format. The rich voter would be willing to pay $1,000 to see this change enacted, while the poor voters would be willing to pay $100 each to prevent it. If each voter's interests are weighted equally, the switch will not be adopted. Yet, in cost-benefit terms, failure to switch results in a net loss of $800.

Under the circumstances, little ingenuity is required to design a proposal that would command unanimous support. The switch could be made conditional, for example, on the rich voter making an additional $500 contribution to the public treasury, which could then be used to reduce the taxes of each poor voter by $250.

Critics may respond that although such transfers would be fine in principle, the poor lack the political muscle to assure they are carried out. In an imperfect world, they argue, we get better results by resolving such issues on a one-person, one-vote basis.

But this response simply will not do. If the poor lack the political power to bargain for compensation in return for supporting a policy that harms them, what gives them the power to block that policy in the first place? But if they have that power, they necessarily have the power to bargain for compensation. After all, any policy that passes the cost-benefit test but creates net losses for the poor can be transformed into a Pareto improvement by simply making the tax system more progressive.

Critics of cost-benefit analysis are correct that using unweighted willingness-to-pay measures virtually assures a mix of public programs that are slanted in favor of the preferences of high-income persons. But rather than abandon cost-benefit analysis, we have a better alternative. We can employ unweighted willingness-to-pay measures without apology, and use the welfare and tax system to compensate low-income families ex ante for the resulting injury. The compensation need not – indeed cannot – occur on a case-by-case basis. Rather, low-income persons could simply be granted the welfare and tax breaks required by distributive justice, plus additional concessions reflecting their expected loss from the implementation of cost-benefit analysis using unweighted willingness-to-pay measures.

My point in offering this defense of standard cost-benefit analysis is not that granting additional political power to the poor would be a bad idea. Rather, it is that abandoning cost-benefit analysis is a gratuitously wasteful way of trying to achieve that goal. Rich and poor alike have an interest in making the economic pie as large as possible. Any policy that passes the cost-benefit test makes the economic pie larger. And when the pie is larger, everyone can have a larger slice.

Measurement Problems

To discover whether an action satisfies the cost-benefit test, we must come up with concrete measures of its costs and benefits. Notwithstanding the logical difficulties raised by claims of incommensurability, this much is clear: constructing plausible measures of the costs and benefits of specific actions is often very difficult. In practice, analysts try to estimate costs and benefits either by using survey methods or by drawing inferences from market behavior. Both approaches, however, are fraught with difficulty.

Survey Methods

How much is the preservation of a virgin redwood forest worth? Proponents of the contingent-valuation method generate estimates by asking people how much they would be willing to pay to see the forest preserved. Responses in such surveys are problematic for several reasons.

One difficulty is that the valuations are often implausibly large. For example, if the amount someone would pay to prevent a specific stretch of coastline from being fouled by an oil spill were applied to all coastlines worldwide, the resulting sum would typically far exceed his total wealth.[6] Responses in contingent-valuation surveys are also highly sensitive to how questions are phrased and to the format provided for responses.[7]

But perhaps the most troubling feature of contingent-valuation surveys is that respondents are often willing to pay more, by several orders of magnitude, to prevent a harmful effect than to undo a harmful effect that has already occurred. Richard Thaler coined the term "loss aversion" to describe this tendency.[8] Loss aversion means not just that the pain of losing a given amount is larger, for most of us, than the pleasure from gaining that same amount. It is much larger.

Thaler illustrates the asymmetry by asking students to consider the following hypothetical questions:

1. By attending class today, you have been exposed to a rare, fatal disease. The probability that you have the disease is one in a thousand. If you have the disease you will die a quick and painless death in one week. There is a cure for the disease that always works, but it has to be taken now. We do not know how much it will cost. You must say now the most you would be willing to pay for this cure. If the cure ends up costing more you won't get it. If it costs less, you will pay the stated price, not the maximum you stated. How much will you pay?
2. We are conducting experiments on the same disease for which we need subjects. A subject will just have to expose him or herself to the disease and risk a one-in-a-thousand chance of death. What is the minimum fee you would accept to become such a subject?[9]

In each scenario, respondents are asked, in effect, how much they value a one in 1,000 reduction in the probability of death. But whereas the first scenario asks how much they would pay to eliminate a risk of death to which they have already been exposed, the second asks them how much they would have to be paid before exposing themselves to a similar risk voluntarily. The median responses were approximately $800 for the first question and $100,000 for the second.[10] Similar disparities between willingness to pay and willingness to accept are observed in contingent-valuation surveys that pose environmental questions.[11] Disparities in other domains are typically smaller, though few surveys find willingness-to-pay values that are more than half as large as the corresponding values for willingness to accept.[12] These disparities, needless to say, pose formidable hurdles for analysts who employ contingent-valuation methods.

Hedonic Methods

These and other problems inherent in survey methods have led many analysts to favor hedonic pricing models, which attempt to infer valuations from observable market behavior. In typical applications, analysts estimate the value of noise reduction by examining how residential housing prices vary with ambient noise levels, or the value of safety by examining how wages vary with workplace injury levels.[13]

Hedonic pricing models assume that the wage-safety gradient tells us how much workers value safety. Is this a tenable assumption? The argument in support of it is a simple application of invisible-hand theory. If an amenity – say, a guardrail on a lathe – costs $50 per month to install and maintain, and if workers value it at $100 per month, then firms that do not install one risk losing valued employees to a competitor who does. After all, if a competitor were to pay a worker $60 per month less than he earns from his current employer, it could cover the cost of the safety device with $10 to spare, while providing an overall compensation package that is $40 per month more attractive than his current employer's.

To this argument, critics respond that labor markets are not workably competitive in practice. Incomplete information, worker immobility, and other imperfections force workers to accept whatever conditions employers offer. But even if a firm were the only employer in a labor market, it would still have a clear incentive to install a $50 safety device that is worth $100 to the worker. Failure to do so would leave cash on the table.

Other critics suggest that workers often do not know about the safety devices they lack. But this claim is also troubling because firms would have strong incentives to call these devices to workers' attention. After all, both the firm and its workers come out ahead when a cost-effective safety device is adopted.

With respect to the charge that labor markets are not effectively competitive, critics of hedonic pricing models have failed to meet the burden of proof. Worker mobility between firms is high, as is entry by new firms into existing markets, and cartel agreements have always been notoriously unstable. Information is never perfect, but if a new employer in town is offering a better deal, word sooner or later gets around.

If, despite these checks, some firms still managed to exploit their workers by paying less than a competitive wage, we should expect these firms to have relatively high profits. In fact, however, we observe just the opposite correlation. Year in and year out, the firms paying the highest wages are most profitable.[14]

But even if labor markets are workably competitive, the same theory of revealed preference that makes hedonic models so attractive also sounds a cautionary note. It calls our attention to a related form of behavioral evidence, namely, the laws we choose to adopt. Scholars in the law and economics movement have long argued that laws tend to evolve in ways that maximize wealth.[15] This characterization presumably also applies to laws regulating health and safety in the workplace, which by now have been enacted by virtually all industrial democracies. These laws pose a challenge to the hedonic pricing model's assumption that safety risks are fully reflected in compensating wage differentials. If this assumption were correct, safety regulations would entail costs that exceed their benefits and therefore should not have been enacted in the first place. But although these regulations have often been criticized on practical grounds, they appear in no imminent political danger.

Does the political success of safety regulation suggest that hedonic pricing models are misleading? I believe it does, but not for the reasons usually given. In what follows I construct an example to illustrate an alternative rationale for safety regulation, one that is independent of market power and imperfect information.

Positional Concerns and Revealed Preference

Consider a hypothetical community with only two members, Sherwin and Gary. Each gets satisfaction from three things – from his income, from his safety on the job, and from his position on the economic ladder. Each must choose between two jobs – a safe job that pays \$300 per week and a risky job that pays \$350 per week. The value of safety to each is \$100 per week, and each evaluates relative income as follows: Having more income than his neighbor provides the equivalent of \$100 per week worth of additional

		Sherwin	
		Safe job @ $300/week	Unsafe job @ $350/week
Gary	Safe job @ $300/week	$400/week each	$300/week for Gary $450/week for Sherwin
	Unsafe job @ $350/week	$450/week for Gary $300/week for Sherwin	$350/week each

Figure 14.1. The effect of concerns about relative income on worker choices regarding safety.

satisfaction; having less income than his neighbor means the equivalent of a $100 per week reduction in satisfaction; and having the same income as his neighbor means no change in the underlying level of satisfaction. Will Sherwin and Gary choose optimally between the two jobs?

If we viewed each person's decision in isolation, the uniquely correct choice would be the safe job. Although it pays $50 per week less than the risky job, the extra safety it provides is worth $100 per week. So if we abstract from the issue of concern about relative income, the value of the safe job is $400 per week (its $300 salary plus $100 worth of safety), which is $50 per week more than the $350 value of the risky job.

Once we incorporate concerns about relative income, however, the decision logic changes in a fundamental way. Now the attractiveness of each choice depends on the job chosen by the other. The four possible combinations of choices and the corresponding levels of satisfaction are shown in Figure 14.1.

Suppose, for example, that Gary chooses the safe job. If Sherwin then chooses the unsafe job, he ends up with total satisfaction worth $450 – $350 in salary plus $100 from having more income than Gary. Gary, for his part, ends up with only $300 worth of total satisfaction – $300 in salary plus $100 from safety minus $100 from having lower income than Sherwin. Alternatively, suppose Gary chooses the unsafe job. Then Sherwin again does better to accept the unsafe job, for by so doing he gets $350 worth of satisfaction rather than only $300. Since the payoff matrix is symmetric, each player's dominant strategy is to choose the unsafe job. Analysts equipped with the hedonic pricing model will conclude that these workers must value the extra safety at less than $50 per week.

But this inference is clearly wrong. Note that if each chooses a safe job, each will get $400 worth of total satisfaction – $300 of income, $100 worth of satisfaction from safety, and zero satisfaction from relative position. If each had instead chosen the unsafe job, each would have had $350 of income, zero satisfaction from safety, and each would again have had the same level of income, so again zero satisfaction from relative position. If we compare the upper-left cell of Figure 14.1 to the lower-right cell, then, we can say unequivocally that Sherwin and Gary would be happier if each took a safe job at lower income than if each chose an unsafe job with more income. By assumption, the extra safety is worth more than its cost.

The discrepancy arises because the job safety choice confronts workers with a Prisoner's Dilemma. If they could choose collectively, they would pick the safe job, an outcome they prefer to what happens when they choose independently. On this interpretation, safety regulation is attractive not because it prevents exploitation, but because it mitigates the consequences of consumption externalities.

Many modern disciples of Adam Smith appear reluctant to introduce concerns about relative position into normative economic models. Yet as Smith himself recognized, such concerns are a basic component of human nature:

> Consumable commodities are either necessaries or luxuries. By necessaries I understand not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even of the lowest order, to be without. A linen shirt, for example, is, strictly speaking, not a necessary of life. The Greeks and Romans lived, I suppose, very comfortably though they had no linen. But in the present times, through the greater part of Europe, a creditable day-labourer would be ashamed to appear in public without a linen shirt, the want of which would be supposed to denote that disgraceful degree of poverty which, it is presumed, nobody can well fall into without extreme bad conduct. Custom, in the same manner, has rendered leather shoes a necessary of life in England. The poorest creditable person of either sex would be ashamed to appear in public without them.[16]

As Smith clearly understood, concerns about relative income need not entail a desire to have more or better goods than one's neighbors. People with low relative income experience not just psychological discomfort but also more tangible economic costs.[17] A resident of a remote Indian mountain village has no need for a car, but a resident of Los Angeles cannot meet even the most minimal demands of social existence without one. A family that wants to send its children to a good school must buy a house in a good school district, yet such houses are often beyond reach for families with low relative income. Similarly, if only 10 percent of houses have views and everybody cares equally strongly about having a view, then only people in the top 10 percent of the income distribution will get one.

Measuring the social value of a consumption good by summing what individuals spend on it is similar to measuring the social value of military armaments by summing the amounts that individual nations spend on them. Both measurements are problematic because they ignore the influence of context on demand.

Consider a simple model in which individuals apportion their income between consumption (C) and workplace safety (S) and in which the representative individual's utility depends not only on her absolute levels of consumption and safety, but also on her relative consumption. For example, suppose the ith individual's utility is given by[18]

$$U_i = U_i[C_i, S_i, R(C_i)], \tag{14.1}$$

where $R(C_i)$ denotes her rank in the consumption distribution, $0 \leq R(C_i) \leq 1$. If $f(C)$ is the density function for the observed values of consumption in the population, then

$$R(C_i) = \int_0^{C_i} f(C)\, dc.$$

Let M_i denote the individual's income, P_c the price of the consumption good, and P_s the price of safety. If the individual takes $f(C)$ as given, the first-order condition for maximum utility is given by

$$U_{i1}/U_{i2} + [U_{i3}\, f(C_i)C]/U_{i2} = P_c/P_s, \tag{14.2}$$

where U_{ij} denotes the first partial derivative of U_i with respect to its jth argument.

The second term on the left-hand side of equation (2) reflects the fact that when an individual buys an additional unit of the consumption good, her payoff is not just the direct utility it provides but also the utility from the implied advance in the consumption ranking. But other individuals also perceive this second reward, and when all respond to it, the resulting consumption ranking remains as before. As a result, consumers spend more on consumption and less on safety than is socially optimal.

Suppose consumers could agree collectively to ignore the effect of individual consumption changes on consumption rank – that is, suppose they could agree to assume that $R'(C) = f(C) = 0$. The first-order condition in equation (2) would then simplify to

$$U_{i1}/U_{i2} = P_c/P_s, \tag{14.3}$$

which is the familiar first-order condition from models in which consumption rank does not matter. Suppressing the rank term would lead individuals to consume less and spend more on safety than before. Equation (3), not equation (2), defines the socially optimal allocation.

The driving force behind this market failure is that the utility from consumption is more context dependent than the utility from safety. If utility had been equally context dependent for each good, there would have been no distortion.

Is the extent to which satisfaction depends on context different in different domains? Sara Solnick and David Hemenway recently conducted a survey of graduate students in the public health program at Harvard University in an attempt to answer this question.[19] They began by asking each subject to choose between the following hypothetical worlds:

A: You earn \$50,000 a year, others earn \$25,000;
B: You earn \$100,000 a year, others earn \$200,000.

Fifty-six percent of subjects chose the first world. Solnick and Hemenway then asked each subject to choose between worlds in which their relative and absolute income levels were the same, but their relative and absolute vacation times differed:

C: You have 2 weeks of vacation each year, others have 1 week;
D: You have 4 weeks of vacation each year, others have 8 weeks.

This time only 20 percent chose the first world, less than half as many as in the first question. On its face, this suggests that satisfaction from consumption is more strongly context dependent than satisfaction from vacation time.

Other important consumption categories also appear to be less sensitive than material goods consumption to interpersonal comparisons. Consider traffic congestion, whose adverse effects on health and psychological well-being are similar to those of prolonged exposure to loud, unpredictable noise.[20] The effect of such noise on subjects in the laboratory occurs independently of the amount of noise to which other subjects are exposed, suggesting that the demand for goods is more context sensitive than the demand for such environmental amenities as freedom from noise and traffic congestion.

Interpersonal comparisons also appear relatively unimportant for savings, at least in the short run. Thus, whereas most of us know what kinds of houses our friends live in and what kinds of cars they drive, we are much less likely to know how large their savings accounts are. But even if everyone's savings balance were on public display, at least some important individual rewards from current consumption would still depend more on context than those from saving. Many parents, for example, might gladly settle for a diminished standard of living in retirement if by saving less they could meet the payments on a house in a better school district.[21] And the same incentives would lead many parents to accept less safe, more regimented, but better paying, jobs. As before, however, the positional gains enjoyed by families that make such choices are offset by the corresponding positional losses experienced by other families.

How might a cost-benefit analyst adjust conventional estimates to counteract the biases introduced by concerns about relative consumption? One simple method would make use of surveys in which subjects are periodically asked to report how much additional income a family would need to maintain a constant level of subjective well-being in the face of a rise in the incomes of others. Using data collected in several European countries, B. M. S. van Praag and Arie Kapteyn estimate an elasticity of roughly 0.6 – that is, that a family would need about a 6 percent increase in its real income to compensate for a 10 percent increase in the incomes of all others in the community.[22] If we take this estimate at face value for illustrative purposes, we can employ it to construct a simple multiplier for adjusting willingness-to-pay values generated by hedonic pricing models.

Suppose, for example, that a study in which wages were regressed on mortality rates in the workplace found that individual workers are willing to give up 2 percent of their incomes each year in exchange for a one in 1,000 reduction in the probability of dying in a workplace accident. This estimate tells us that a worker earning $50,000 per year would be would be willing to pay $1,000 per year for the additional safety, even though the expenditure would reduce his relative consumption by 2 percent. The Kapteyn–van Praag estimate suggests that this worker would be willing to pay roughly $600 more for the same increment in safety if he could be assured that his relative income would be unaffected by the expenditure – as would be the case, for example, if everyone else made similar expenditures on safety.

An adjustment based on the van Praag–Kapteyn survey data would thus call an upward revision by 60 percent in the willingness-to-pay values inferred from hedonic pricing models. It would be easy to quarrel, of course, with an adjustment procedure based on survey responses like these. Other, more objective procedures might be pursued. Elsewhere, for example, I have argued that one can infer the value of relative income by examining the relationship between wages, local rank, and productivity among groups of coworkers.[23] In any event, the mere fact that an adjustment procedure may be flawed clearly does not imply that it yields worse estimates than we would get by simply ignoring concerns about relative consumption.

In sum, if demands for some goods are more highly context sensitive than demands for others, then individual spending decisions cannot be aggregated to estimate social valuations for cost-benefit analysis. In general, the sum of individual valuations will be smaller than social value for goods whose demands are relatively sensitive to context and greater than social value for those whose demands are relatively insensitive to context. And because contextual forces influence demands in powerful ways,[24] we have ample reason to be skeptical of hedonic pricing models, even those based on perfectly competitive markets with complete information.

As before, however, the implication is not that the cost-benefit approach is invalid as a matter of principle. Rather, it is that, as currently implemented, its prescriptions may be substantially misleading. If so, the remedy is not to abandon cost-benefit analysis but to amend conventional estimating procedures.

Impulse-Control Problems and Revealed Preference

Hedonic pricing models also assume that we can infer the values people place on future events by observing the choices they make. On this view, if a person accepts a one in 10 chance of contracting a serious illness 1 year from now in return for a payment of $100 now, then the cost of taking that risk, expressed as a present value, cannot be more than $100. Compelling experimental evidence, however, suggests grounds for skepticism.[25] Consider, for example, the pair of choices A and B:

A: $100 tomorrow versus $105 a week from tomorrow;
B: $100 after 52 weeks versus $105 after 53 weeks.

The rational choice model on which hedonic pricing models are based says that people will discount future costs and benefits exponentially at their respective rates of time preference. If so, people should always choose similarly under alternatives A and B. Since the larger payoff comes a week later in each case, the ordering of the present values of the two alternatives must be the same in both, irrespective of the rate at which people discount. When people confront such choices in practice, however, most pick the $100 option in A, whereas most choose the $105 option in B.

Substantial experimental evidence suggests that individuals discount future costs and benefits not exponentially, as assumed by the rational choice model, but hyperbolically.[26] The psychological impact of a cost or benefit falls much more sharply with delay under hyperbolic discounting than under exponential discounting. One consequence is that preference reversals of the kind just discussed are all but inevitable under hyperbolic discounting. The classic reversal involves choosing the larger, later reward when both alternatives occur with substantial delay, then switching to the smaller, earlier reward when its delay falls below some threshold. Thus, from the pair of alternatives labeled B above, in which both rewards come only after a relatively long delay, most subjects chose the larger, later reward, whereas from the pair labeled A, most chose the earlier, smaller reward.

The tendency to discount future costs and benefits hyperbolically gives rise to a variety of familiar impulse-control problems and, in turn, to a variety of strategies for solving them. Anticipating their temptation to overeat, people often try to limit the quantities of sweets, salted nuts, and other delicacies they keep on hand. Anticipating their temptation to spend cash in their checking accounts, people enroll in payroll deduction savings plans. Foreseeing the difficulty of putting down a good mystery novel in midstream, many people know better than to start one on the evening before an important meeting. Reformed smokers seek the company of nonsmokers when they first try to kick the habit and are more likely than others to favor laws that limit smoking in public places. The recovering alcoholic avoids cocktail lounges.

Effective as these bootstrap self-control techniques may often be, they are far from perfect. Many people continue to express regret about having overeaten, having drunk and smoked too much, having saved too little, having stayed up too late, having watched too much television, and so on. The exponential discounting model urges us to dismiss these expressions as sour grapes. But from the perspective of the hyperbolic discounting model, these same expressions are coherent. In each case, the actor chose an inferior option when a better one was available, and later feels genuinely sorry about it.

Hedonic pricing models use observed choices to infer discount rates, which cost-benefit analysts then use to compute present values. To the extent that many important intertemporal choices are driven by hyperbolic discounting, conventional methods will give too little weight to future costs and benefits.

Status Quo Bias

Opposition to cost-benefit analysis may also stem from the fact that the costs of a policy change are often far easier to quantify than its benefits, especially in the domains of environmental policy and health and safety policy. In both fields, consensus about how to measure benefits has proved especially elusive. The upshot is that policy decisions in these arenas tend to be driven primarily by cost considerations, resulting in a bias in favor of the status quo. This bias may help explain why advocates of change are overrepresented among opponents of cost-benefit analysis.

The fact that benefits are more difficult to measure than costs does not provide a compelling reason to abandon cost-benefit analysis, just as the fact that costs are easier to forecast than revenues does not provide a compelling reason for firms to abandon profit maximization. In each case, we do better to act on the best information available than to act on no information at all.

Concluding Remarks

From the preceding discussion, I draw two conclusions. One is that critics have failed to offer persuasive arguments that cost-benefit analysis is objectionable as a matter of principle. The other is that many of the methods used by cost-benefit analysts generate systematically biased prescriptions. Hedonic pricing methods overstate the value of goods and activities whose demands are relatively context sensitive. And they give too much weight to current costs and benefits, too little weight to those that occur in the future. These biases suggest an answer to the question posed in my title. Cost-benefit analysis as currently practiced may be controversial simply because it often generates misleading prescriptions.

I conclude by considering a more speculative explanation for opposition to cost-benefit analysis, one rooted in the distinction between consequentialist and deontological moral theories. The deontologists insist that immutable moral principles distinguish right conduct from wrong conduct, irrespective of costs and benefits. They insist, for example, that stealing is wrong not because it does more harm than good, but simply because it violates the victim's rights. The consequentialist resists such absolute prescriptions, confident that there could always be *some* conditions in which the gains from stealing might outweigh its costs.

Yet even the most committed consequentialists seem to recognize that statements like "Stealing is permissible whenever its benefits exceed its costs" are not rhetorically effective for teaching their children moral values. Indeed, like the deontologists, most consequentialists teach their children that stealing is wrong as a matter of principle. Elsewhere I have argued that once we acknowledge the strategic role of moral emotions in solving commitment problems, this posture is coherent, even in purely consequentialist terms.[27]

Yet a potentially more worrisome aspect of the consequentialist position remains, which is that people who view their ethical choices in cost-benefit terms must also construct their own estimates of the relevant costs and benefits. The obvious concern is that their estimates will be self-serving. More than 90 percent of all drivers, for example, feel sure they are better than average.[28] More than 99 percent of high-school students think they are above average in terms of their ability to get along with others.[29] Ninety-four percent of college professors believe they are more productive than their average colleague.[30] The same forces that make us overestimate our skills can be expected also to distort the estimates that underlie our ethical judgments. And if these self-serving calculations lead some to disregard the common good, their example will make others more apt to do likewise.

Needless to say, people may also be prone to self-serving biases in their interpretations of deontological moral principles. In the end, which approach entails the greater risk is an empirical question. But it is at least possible that consequentialist thinking could lead to a worse outcome on balance. If this were shown to be so, consequentialists would have little choice but to endorse the deontological position (much as an atheist might support fundamentalist religious institutions on the view that threats of hell-fire and damnation are the only practical way to get people to behave themselves). They would have to view cost-benefit analysis as correct in principle yet best avoided in practice.

I hasten to add that critics of cost-benefit analysis have made no such showing. And unless they do, it seems certain that cost-benefit analysis will continue to play an important role in decision making. Under the circumstances, both friends and foes of cost-benefit analysis have a shared interest in trying to eliminate the biases that distort its prescriptions.

Notes

1. For an overview, see Robert Kuttner, Everything for Sale (1997).
2. The few who did object would likely invoke a variation of the "slippery-slope" argument, which holds that allowing even a single small step will lead to an inevitable slide to the bottom. Yet we move partway down slippery slopes all the time, as when we amend the laws of free speech to prohibit people from yelling "fire" in a crowded theater in which there is no fire.

3. See, for example, Steven Kelman, An Ethical Critique of Cost-Benefit Analysis, 5 Regulation 33 (1981).
4. See, for example, John Jamieson, Carswell Smart, & Bernard Williams, Utilitarianism: For and Against (1973).
5. Stanley Jevons, The Theory of Political Economy 72–73 (1941) (1871).
6. See I. Ritov & Daniel Kahneman, How People Value the Environment: Attitudes vs. Economic Values, in Psychological Approaches to Environmental and Ethical Issues in Management 33–51 (M. Bazerman *et al.* eds. 1997); and Daniel Kahneman & Jack Knetsch, Valuing Public Goods: The Purchase of Moral Satisfaction, 22 J. Envtl. Econ. & Mgmt. 57 (1992).
7. William H. Desvousges, John W. Payne, & David A. Schkade, How People Respond to Contingent Valuation Questions (EPA Grant No. R824310 Final Report, April 1998).
8. Richard Thaler, Toward a Positive Theory of Consumer Choice, 1 J. Econ. Behav. & Org. 39 (1980).
9. Richard Thaler, Precommitment and the Value of a Life, in The Value of Life and Safety 178–79 (M. W. Jones-Lee ed. 1982).
10. *Id.* at 179.
11. Ritov & Kahneman, *supra* note 6.
12. Rebecca Boyce *et al.*, An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity, 82 Am. Econ. Rev. 1366 (1992).
13. See, for example, Richard Thaler & Sherwin Rosen, The Value of Saving a Life: Evidence from the Labor Market, in Household Production and Consumption 265 (N. Terlekyj ed. 1976).
14. See Lawrence Seidman, The Return of the Profit Rate to the Wage Equation, 61 Rev. Econ. & Stat. 139 (1979), and numerous studies cited therein.
15. See, for example, Richard A. Posner, Economic Analysis of Law (5th ed. 1998).
16. Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations, bk. 5, ch. II. pt. II, art. 4 (1952) (1776).
17. On this point, see especially Amartya K. Sen, The Standard of Living (1989).
18. For a more detailed discussion of the model that follows, see Robert H. Frank, The Demand for Unobservable and Other Nonpositional Goods, 75 Am. Econ. Rev. 101 (1985).
19. Sara J. Solnick & David Hemenway, Is More Always Better? A Survey on Positional Concerns, 37 J. Econ. Behav. & Org. 373 (1998).
20. For a survey of the relevant studies, see Robert H. Frank, Luxury Fever, ch. 6 (1999).
21. Some object that a desire for high consumption rank cannot really explain low savings rates, since those who save too little now simply consign themselves to having low consumption rank in the future. Yet, as noted, having lower consumption rank in the future may be an acceptable price to pay for the ability to have high rank with respect to some forms of current consumption. What is more, to the extent that driving the right cars and wearing the right clothes function as signals of ability, and thereby help people land better jobs or more lucrative contracts, low savings now may not even entail reduced consumption rank in the future. But whereas this may be true from the perspective of each individual, it is surely not true for society as a whole. For when all of us

spend more to signal our abilities, the relative strength of each signal remains unchanged.

22. B. M. S. van Praag & Arie Kapteyn, Further Evidence on the Individual Welfare Function of Income, 4 Eur. Econ. Rev. 33 (1973).
23. Robert H. Frank, Are Workers Paid Their Marginal Products? 74 Am. Econ. Rev. 549 (1984).
24. See Robert H. Frank, Choosing the Right Pond (1985); and Frank, *supra* note 20.
25. See, for example, the papers in Choice over Time (Jon Elster & George Loewenstein eds. 1993).
26. For detailed summary of the relevant evidence, see George Ainslie, Picoeconomics (1992).
27. See Robert H. Frank, Passions within Reason (1988). For a related discussion, see Eric A. Posner, The Strategic Basis of Unprincipled Behavior: A Critique of the Incommensurability Thesis, 146 U. Pa. L. Rev. 1185 (1998).
28. See Thomas Gilovich, How We Know What Isn't So (1991).
29. College Board, Student Descriptive Questionnaire (1976–77).
30. P. Cross, Not *Can* but *Will* College Teaching Be Improved? New Directions Higher Educ., Spring 1977, at 1.

FIFTEEN

Capability and Well-Being

Amartya Sen*

Amartya Sen (1933–) was born and educated in India before completing his doctorate in economics at Cambridge University. He has taught in India, England, and the United States and is currently the Lamont University Professor at Harvard University. He is one of the most widely read and influential living economists. His books have been translated into more than thirty languages. In 1998, he was awarded the Nobel Price in Economics for his work on welfare economics, poverty and famines, and human development. He has also made major contributions to contemporary political philosophy. In this essay, he proposes that alternatives be appraised by looking to the capabilities they provide for individuals rather than only by individual utilities, incomes, or resources (as in commonly used theories).

1. Introduction

Capability is not an awfully attractive word. It has a technocratic sound, and to some it might even suggest the image of nuclear war strategists rubbing their hands in pleasure over some contingent plan of heroic barbarity. The term is not much redeemed by the historical Capability Brown praising particular pieces of *land* – not human beings – on the solid real-estate ground that they 'had capabilities'. Perhaps a nicer word could have been chosen when some years ago I tried to explore a particular approach to well-being and advantage in terms of a person's ability to do valuable acts or reach valuable states of being.[1] The expression was picked to represent the alternative

* For helpful discussions, I am most grateful to G. A. Cohen, Partha Dasgupta, Jean Drèze, Hilary Putnam, Ruth Anna Putnam, Martha Nussbaum, Derek Parfit, John Rawls, John Roemer, and Thomas Scanlon.

Pages 30–42 and 46–53 of *The Quality of Life*, edited by Amartya Sen and Martha Nussbaum. Published by Oxford University Press, 1993. With the author's permission, section 9 of the essay is omitted. That section comments on an essay by G. A. Cohen that appeared in the same volume.

combinations of things a person is able to do or be – the various 'functionings' he or she can achieve.[2]

The capability approach to a person's advantage is concerned with evaluating it in terms of his or her actual ability to achieve various valuable functionings as a part of living. The corresponding approach to social advantage – for aggregative appraisal as well as for the choice of institutions and policy – takes the sets of individual capabilities as constituting an indispensable and central part of the relevant informational base of such evaluation. It differs from other approaches using other informational focuses, for example, personal utility (focusing on pleasures, happiness, or desire fulfilment), absolute or relative opulence (focusing on commodity bundles, real income, or real wealth), assessments of negative freedoms (focusing on procedural fulfilment of libertarian rights and rules of non-interference), comparisons of means of freedom (e.g. focusing on the holdings of 'primary goods', as in the Rawlsian theory of justice), and comparisons of resource holdings as a basis of just equality (e.g. as in Dworkin's criterion of 'equality of resources').

Different aspects of the capability approach have been discussed, extended, used, or criticized by several authors, and as a result the advantages and difficulties of the approach have become more transparent.[3] There is, however, a need for a clearer and more connected account of the whole approach, particularly in view of some interpretational problems that have arisen in its assessment and use. This paper is an attempt at a clarificatory analysis at an elementary level. I shall also try to respond briefly to some interesting criticisms that have been made.

2. Functionings, Capability, and Values

Perhaps the most primitive notion in this approach concerns 'functionings'. *Functionings* represent parts of the state of a person – in particular the various things that he or she manages to do or be in leading a life. The *capability* of a person reflects the alternative combinations of functionings the person can achieve, and from which he or she can choose one collection.[4] The approach is based on a view of living as a combination of various 'doings and beings', with quality of life to be assessed in terms of the capability to achieve valuable functionings.

Some functionings are very elementary, such as being adequately nourished, being in good health, etc., and these may be strongly valued by all, for obvious reasons. Others may be more complex, but still widely valued, such as achieving self-respect or being socially integrated. Individuals may, however, differ a good deal from each other in the weights they attach to these different functionings – valuable though they may all be – and the assessment of individual and social advantages must be alive to these variations.

In the context of some types of social analysis, for example, in dealing with extreme poverty in developing economies, we may be able to go a fairly long distance with a relatively small number of centrally important functionings and the corresponding basic capabilities (e.g. the ability to be well nourished and well sheltered, the capability of escaping avoidable morbidity and premature mortality, and so forth). In other contexts, including more general problems of economic development, the list may have to be much longer and much more diverse.

Choices have to be faced in the delineation of the *relevant* functionings. The format always permits additional 'achievements' to be defined and included. Many functionings are of no great interest to the person (e.g. using a *particular* washing powder – much like other washing powders).[5] There is no escape from the problem of evaluation in selecting a class of functionings in the description and appraisal of capabilities. The focus has to be related to the underlying concerns and values, in terms of which some definable functionings may be important and others quite trivial and negligible. The need for selection and discrimination is neither an embarrassment, nor a unique difficulty, for the conceptualization of functioning and capability.

3. Value-Objects and Evaluative Spaces

In an evaluative exercise, we can distinguish between two different questions: (1) *What* are the objects of value? (2) *How valuable* are the respective objects? Even though *formally* the former question is an elementary aspect of the latter (in the sense that the objects of value are those that have positive weights), nevertheless the identification of the objects of value is *substantively* the primary exercise which makes it possible to pursue the second question.

Furthermore, the very identification of the set of value-objects, with positive weights, itself precipitates a 'dominance ranking' (x is at least as high as y if it yields at least as much of *each* of the valued objects). This dominance ranking, which can be shown to have standard regularity properties such as transitivity, can indeed take us some distance – often quite a long distance – in the evaluative exercise.[6]

The identification of the objects of value specifies what may be called an *evaluative space*. In standard utilitarian analysis, for example, the evaluative space consists of the individual utilities (defined in the usual terms of pleasures, happiness, or desire fulfilment). Indeed, a complete evaluative approach entails a class of 'informational constraints' in the form of ruling out *directly* evaluative use of various types of information, to wit, those that do not belong to the evaluative space.[7]

The capability approach is concerned primarily with the identification of value-objects, and sees the evaluative space in terms of functionings and capabilities to function. This is, of course, itself a deeply evaluative exercise, but answering question (1), on the identification of the objects of value, does not, on its own, yield a particular answer to question (2), regarding their relative values. The latter calls for a further evaluative exercise. Various substantive ways of evaluating functionings and capabilities can all belong to the general capability approach.

The selection of the evaluative space has a good deal of cutting power on its own, both because of what it *includes* as potentially valuable and because of what it *excludes*. For example, because of the nature of the evaluative space, the capability approach differs from utilitarian evaluation (more generally 'welfarist' evaluation[8]) in making room for a variety of human acts and states as important in themselves (not just *because* they may produce utility, nor just to the *extent* that they yield utility).[9] It also makes room for valuing various freedoms – in the form of capabilities. On the other side, the approach does not attach direct – as opposed to derivative – importance to the *means* of living or *means* of freedom (e.g. real income, wealth, opulence, primary goods, or resources), as some other approaches do. These variables are not part of the evaluative space, though they can indirectly influence the evaluation through their effects on variables included in that space.

4. Capability and Freedom

The freedom to lead different types of life is reflected in the person's capability set. The capability of a person depends on a variety of factors, including personal characteristics and social arrangements. A full accounting of individual freedom must, of course, go beyond the capabilities of personal living and pay attention to the person's other objectives (e.g. social goals not directly related to one's own life), but human capabilities constitute an important part of individual freedom.

Freedom, of course, is not an unproblematic concept. For example, if we do not have the courage to choose to live in a particular way, even though we *could* live that way if we so chose, can it be said that we do have the freedom to live that way, i.e. the corresponding capability? It is not my purpose here to brush under the carpet difficult questions of this – and other – types. In so far as there are genuine ambiguities in the concept of freedom, that should be reflected in corresponding ambiguities in the characterization of capability. This relates to a methodological point, which I have tried to defend elsewhere, that if an underlying idea has an essential ambiguity, a precise formulation of that idea must try to *capture* that ambiguity rather than hide or eliminate it.[10]

Comparisons of freedom raise interesting issues of evaluation. The claim is sometimes made that freedom must be valued independently of the values and preferences of the person whose freedom is being assessed, since it concerns the 'range' of choice a person has – *not* how she values the elements in that range or what she chooses from it. I do not believe for an instant that this claim is sustainable (despite some superficial plausibility), but had it been correct, it would have been a rather momentous conclusion, driving a wedge between the evaluation of *achievements* and that of *freedoms*. It would, in particular, be then possible to assess the freedom of a person independently of – or prior to – the assessment of the alternatives between which the person can choose.[11]

How can we judge the goodness of a 'range' of choice independently of – or prior to – considering the nature of the alternatives that constitute that range? Some comparisons can, of course, be made in terms of set inclusion, for example, that reducing the 'menu' from which one can choose will *not* increase one's freedom.[12] But whenever neither set is entirely included in the other, we have to go beyond such 'subset reasoning'.

One alternative is simply to *count* the number of elements in the set as reflecting the value of the range of choice.[13] But this number-counting procedure leads to a rather peculiar accounting of freedom. It is odd to conclude that the freedom of a person is no less when she has to choose between three alternatives which she sees respectively as 'bad', 'awful', and 'gruesome' than when she has the choice between three alternatives which she assesses as 'good', 'excellent', and 'superb'.[14] Further, it is always possible to add trivially to the number of options one has (e.g. tearing one's hair, cutting one's ears, slicing one's toes, or jumping through the window), and it would be amazing to see such additions as compensating for the loss of really valued options.[15] The assessment of the elements in a range of choice has to be linked to the evaluation of the freedom to choose among that range.[16]

5. Value-Purposes and Distinct Exercises

While the identification of value-objects and the specification of an evaluative space involve norms, the nature of the norms must depend on precisely what the purpose of the evaluation is. Assessing well-being may take us in one direction; judging achievement in terms of the person's *overall* goals may take us in a somewhat different direction, since a person can have objectives other than the pursuit of his or her own well-being. Judging achievement of either kind may also differ from the evaluation of the *freedom* to achieve, since a person can be advantaged in having more freedom and still end up achieving less.

We can make a fourfold classification of points of evaluative interest in assessing human advantage, based on two different distinctions. One distinction is between (1.1) the promotion of the person's *well-being*, and (1.2) the pursuit of the person's overall *agency goals*. The latter encompasses the goals that a person has reasons to adopt, which can *inter alia* include goals other than the advancement of his or her own well-being. It can thus generate orderings different from that of well-being. The second distinction is between (2.1) *achievement*, and (2.2) the *freedom to achieve*. This contrast can be applied both to the perspective of well-being and to that of agency. The two distinctions together yield four different concepts of advantage, related to a person: (1) 'well-being achievement', (2) 'agency achievement', (3) 'well-being freedom', and (4) 'agency freedom'. These different notions, which I have tried to discuss more extensively elsewhere, are not, of course, unrelated to each other, but nor are they necessarily identical.[17]

The assessment of each of these four types of benefit involves an evaluative exercise, but they are not the *same* evaluative exercise. They can also have very disparate bearings on matters to which the evaluation and comparison of individual advantages are relevant. For example, in determining whether a person is deprived in a way that calls for assistance from others or from the state, a person's well-being may be, arguably, more relevant than his agency success (e.g. the state may have better grounds for offering support to a person for overcoming hunger or illness than for helping him to build a monument to his hero, even if he himself attaches more importance to the monument than to the removal of his hunger or illness). Furthermore, for adult citizens, *well-being freedom* may be more relevant to state policy, in this context, than *well-being achievement* (e.g. the state may have reason to offer a person adequate opportunities to overcome hunger, but not to insist that he must take up that offer and cease to be hungry). Interpersonal comparisons can be of many distinct types, with possibly dissimilar evaluative interests. Despite the interdependences between the different value-purposes, they can generate quite distinct exercises with partly divergent concentration and relevance.

6. Well-Being, Agency, and Living Standards

The well-being achievement of a person can be seen as an evaluation of the 'well-ness' of the person's state of being (rather than, say, the goodness of her contribution to the country, or her success in achieving her overall goals). The exercise, then, is that of assessing the constituent elements of the person's being seen from the perspective of her own personal welfare. The different functionings of the person will make up these constituent elements.

This does not, of course, imply that a person's well-being cannot be 'other-regarding'. Rather, the effect of 'other-regarding' concerns on one's well-being has to operate *through* some feature of the person's own being. Doing good may make a person contented or fulfilled, and these are functioning achievements of importance. In this approach, functionings are seen as central to the *nature* of well-being, even though the *sources* of well-being could easily be external to the person.

The functionings relevant for well-being vary from such elementary ones as escaping morbidity and mortality, being adequately nourished, having mobility, etc., to complex ones such as being happy, achieving self-respect, taking part in the life of the community, appearing in public without shame (the last a functioning that was illuminatingly discussed by Adam Smith[18]). The claim is that the functionings make up a person's being, and the evaluation of a person's well-being has to take the form of an assessment of these constituent elements.

If the value-purpose is changed from checking the 'well-ness' of the person's being to assessing the person's success in the pursuit of all the objectives that he has reason to promote, then the exercise becomes one of evaluation of 'agency achievement', rather than of well-being achievement. For this exercise, the space of functionings may be rather restrictive, since the person's goals may well include other types of objective (going well beyond the person's own state of being). Also, the difference between agency achievement and well-being achievement is not only a matter of *space* (the former taking us beyond the person's own life and functionings), but also one of differential *weighting* of the shared elements (i.e. for the functionings that are pertinent both to one's well-being and to one's other objectives, possibly different weights may be attached in agency evaluation *vis-à-vis* well-being appraisal).

The assessment of agency success is a broader exercise than the evaluation of well-being. It is also possible to consider 'narrower' exercises than the appraisal of well-being. A particularly important one is that of evaluating a person's *standard of living*. This, too, may take the form of focusing on the person's functionings, but in this case we may have to concentrate only on those influences on well-being that come from the nature of his *own* life, rather than from 'other-regarding' objectives or impersonal concerns. For example, the happiness generated by a purely other-regarding achievement (e.g. the freeing of political prisoners in distant countries) may enhance the person's well-being without, in any obvious sense, raising his living standard.

In the ethical context, the explicit recognition that one's well-being may often be affected by the nature of other people's lives is not, of course, new. Even Emperor Asoka, in the third century BC, noted the distinction clearly in one of his famous 'rock edicts' in the process of defining what should count as an injury to a person: 'And, if misfortune befalls the friends, acquaintances, companions and relations of persons who are full of affection [towards the former], even though they are themselves well provided for, [this misfortune] is also an injury to their own selves.'[19] The inability to be happy, which will be widely recognized as a failure of an important functioning (even though not the *only* important one, except in the hedonist version of utilitarianism), may arise either from sources within one's own life (e.g. being ill, or undernourished, or otherwise deprived), or from sources outside it (e.g. the pain that comes from sympathizing with others' misery). While both types of factor affect one's well-being, the case for excluding the latter from the assessment, specifically, of one's living standards would seem fairly reasonable, since the latter relates primarily to the lives of others, rather than one's own.[20]

7. Why Capability, Not Just Achievement?

The preceding discussion on the achievement of well-being and living standards has been related to functionings rather than to capabilities. This was done by design to introduce distinct problems in sequence, even though eventually an integrated view will have to be taken. In fact, the capability approach, as the terminology indicates, sees the capability set as the primary informational base. Why should we have to broaden our attention from functionings to capability?

We should first note that capabilities are defined derivatively from functionings. In the space of functionings any point, representing an *n*-tuple of functionings, reflects a combination of the person's doings and beings, relevant to the exercise. The capability is a *set* of such functioning *n*-tuples, representing the various alternative combinations of beings and doings any one (combination) of which the person can choose.[21] Capability is thus defined in the *space* of functionings. If a functioning achievement (in the form of an n-tuple of functionings) is a *point* in that space, capability is a *set* of such points (representing the alternative functioning n-tuples from which one n-tuple can be chosen).

Note further that the capability set contains information about the actual functioning n-tuple chosen, since it too is obviously among the feasible n-tuples. The evaluation of a capability set may be based on the assessment of the particular n-tuple chosen from that set. Evaluation according to the achieved functioning combination is thus a 'special case' of evaluation on the basis of the capability set as a whole. In this sense, well-being achievement can be assessed on the basis of the capability set, even when no freedom-type notion influences that achievement. In this case, in evaluating the capability set for the value-purpose of assessing well-being achievement, we would simply have to identify the value of the capability set with the value of the achieved functioning n-tuple in it. The procedure of equating the value of the capability set to the value of *one* of the elements of that set has been called 'elementary evaluation'.[22]

Clearly, there is *at least* no informational loss in seeing well-being evaluation in terms of capabilities, rather than directly in terms of the achieved, or chosen, or maximal functioning n-tuple. While this indicates that the informational base of capability is at least as adequate as that of achieved functionings, the claim in favour of the capability perspective is, in fact, stronger. The advantages of the extension arise from two rather different types of consideration.

First, we may be interested not merely in examining 'well-being achievement', but also 'well-being freedom'. A person's actual freedom to live well and be well is of some interest in social as well as personal evaluation.[23] Even if we were to take the view, which will be disputed presently, that well-being achievement depends only on the achieved functionings, the 'well-being freedom' of a person will represent the freedom to enjoy the various possible well-beings associated with the different functioning n-tuples in the capability set.[24]

Second, freedom may have intrinsic importance for the person's well-being achievement. Acting freely and being able to choose may be directly conducive to well-being, not just because more freedom may make better alternatives available. This view is contrary to the one typically assumed in standard consumer theory, in which the contribution of a set of feasible choices is judged exclusively by the value of the best element available.[25] Even the removal of all the elements of a feasible set (e.g. of a 'budget set') other than the chosen best element is seen, in that theory, as no real loss, since the freedom to choose does not, in this view, matter in itself.

In contrast, if choosing is seen as a part of living (and 'doing x' is distinguished from 'choosing to do x and doing it'), then even 'well-being achievement' need not be independent of the freedom reflected in the capability set.[26] In that case, both 'well-being achievement' and 'well-being freedom' will have to be assessed in terms of capability sets. Both must then involve 'set evaluation' in a non-elementary way (i.e. without limiting the usable informational content of capability sets through elementary evaluation).

There are many formal problems involved in the evaluation of freedom and the relationship between freedom and achievement.[27] It is, in fact, possible to characterize functionings in a 'refined' way to take note of the 'counterfactual' opportunities, so that the characteristic of relating well-being achievement to functioning n-tuples could be retained without losing the substantive connection of well-being achievement to the freedom of choice enjoyed by the person. Corresponding to the functioning x, a 'refined' functioning (x/S) takes the form of 'having functioning x through choosing it from the set S'.[28]

Sometimes even our ordinary language presents functionings in a refined way. For example, fasting is not just starving, but starving through rejecting the option of eating. The distinction is obviously important in many social contexts: we may, for example, try to eliminate involuntary hunger, but not wish to forbid fasting. The importance of seeing functionings in a refined way relates to the relevance of choice in our lives. The role of the choice involved in a capability set has been discussed above in the context of well-being only, but similar arguments apply to the assessment of agency achievement and the standard of living.[29]

8. Basic Capability and Poverty

For some evaluative exercises, it may be useful to identify a subset of crucially important capabilities dealing with what have come to be known as 'basic needs'.[30] There tends to be a fair amount of agreement on the extreme urgency of a class of needs. Particular moral and political importance may well be attached to fulfilling well-recognized, urgent claims.[31]

It is possible to argue that equality in the fulfilment of certain 'basic capabilities' provides an especially plausible approach to egalitarianism in the presence of elementary deprivation.[32] The term 'basic capabilities', used in Sen (1980), was intended to separate out the ability to satisfy certain crucially important functionings up to certain minimally adequate levels. The identification of minimally acceptable levels of certain basic capabilities (below which people count as being scandalously 'deprived') can provide a possible approach to poverty, and I shall comment on the relation of this strategy to more traditional income-focused analyses of poverty. But it is also important to recognize that the use of the capability approach is not confined to basic capabilities only.[33]

Turning to poverty analysis, identifying a minimal combination of basic capabilities can be a good way of setting up the problem of diagnosing and measuring poverty. It can lead to results quite different from those obtained by concentrating on the inadequacy of income as the criterion of identifying the poor.[34] The conversion of income into basic capabilities may vary greatly between individuals and also between different societies, so that the ability to reach minimally acceptable levels of basic capabilities can go with varying levels of minimally adequate incomes. The income-centred view of poverty, based on specifying an interpersonally invariant 'poverty line' income, may be very misleading in the identification and evaluation of poverty.

However, the point is sometimes made that poverty must, in some sense, be a matter of inadequacy of income, rather than a failure of capabilities, and this might suggest that the capability approach to poverty is 'essentially wrong-headed'. This objection overlooks both the motivational underpinning of poverty analysis and the close correspondence between capability failure and income inadequacy when the latter is defined taking note of *parametric variations* in income-capability relations.

Since income is not desired for its own sake, any income-based notion of poverty must refer – directly or indirectly – to those basic ends which are promoted by income as means. Indeed, in poverty studies related to less developed countries, the 'poverty line' income is often derived explicitly with reference to nutritional norms. Once it is recognized that the relation between income and capabilities varies between communities and between people in the same community, the minimally adequate income level for reaching the same minimally acceptable capability levels will be seen as variable – depending on personal and social characteristics. However, as long as minimal capabilities can be achieved by enhancing the income level (given the other personal and social characteristics on which capabilities depend), it will be possible (for the specified personal and social characteristics) to identify the minimally adequate income for reaching the minimally acceptable capability levels. Once this correspondence is established, it would not really matter whether poverty is defined in terms of a failure of basic capability or as a failure to have the *corresponding* minimally adequate income.[35]

Thus, the motivationally more accurate characterization of poverty as a failure of basic capabilities can also be seen in the more traditional format of an income inadequacy. The difference in formulation is unimportant. What is really important is to take note of the interpersonal and intersocial variations in the relation between incomes and capabilities. That is where the distinctive contribution of the capability approach to poverty analysis lies.

10. The Aristotelian Connections and Contrasts

In earlier writings I have commented on the connection of the capability approach with some arguments used by Adam Smith and Karl Marx.[36] However, the most powerful conceptual connections would appear to be with the Aristotelian view of the human good. Martha Nussbaum (1988, 1990) has discussed illuminatingly the Aristotelian analysis of 'political distribution', and its relation to the capability approach. The Aristotelian account of the human good is explicitly linked with the necessity to 'first ascertain the function of man' and it then proceeds to explore 'life in the sense of activity'.[37] The basis of a fair distribution of capability to function is given a central place in the Aristotelian theory of political distribution. In interpreting Aristotle's extensive writings on ethics and politics, it is possible to note some ambiguity and indeed to find some tension between different propositions presented by him, but his recognition of the crucial importance of a person's functionings and capabilities seems to emerge clearly enough, especially in the political context of distributive arrangements.

While the Aristotelian link is undoubtedly important, it should also be noted that there are some substantial differences between the way functionings and capabilities are used in what I have been calling the capability approach and the way they are dealt with in Aristotle's own analysis. Aristotle believes, as Nussbaum (1988) notes, 'that there is just one list of functionings (at least at a certain level of generality) that do in fact constitute human good living' (p. 152). That view would not be inconsistent with the capability approach presented here, but *not*, by any means, *required* by it.

The capability approach has indeed been used (for example, in Sen, 1983*c*, 1984) to argue that while the *commodity* requirements of such capabilities as 'being able to take part in the life of the community' or 'being able to appear in public without shame' vary greatly from one community to another (thereby giving the 'poverty line' a relativist character in the space of commodities), there is much less variation in the *capabilities* that are aimed at through the use of these commodities. This argument, suggesting less variability at a more intrinsic level, has clear links with Aristotle's identification of 'non-relative virtues', but the Aristotelian claims of uniqueness go much further.[38]

Martha Nussbaum, as an Aristotelian, notes this distinction, and also points to Aristotle's robust use of an objectivist framework based on a particular reading of human nature. She suggests the following:

> It seems to me, then, that Sen needs to be more radical than he has been so far in his criticism of the utilitarian accounts of well-being, by introducing an objective normative account of human functioning and by describing a procedure of objective evaluation by which functionings can be assessed for their contribution to the good human life.[39]

I accept that this would indeed be a systematic way of eliminating the incompleteness of the capability approach. I certainly have no great objection to anyone going on that route. My difficulty with accepting that as the *only* route on which to travel arises partly from the concern that this view of human nature (with a unique list of functionings for a good human life) may be tremendously over-specified, and also from my inclination to argue about the nature and importance of the type of objectivity involved in this approach. But mostly my intransigence arises, in fact, from the consideration that the use of the capability approach as such does not require taking that route, and the deliberate incompleteness of the capability approach permits other routes to be taken which also have some plausibility. It is, in fact, the feasibility as well as the usefulness of a general approach (to be distinguished from a complete evaluative blueprint) that seems to me to provide good grounds for separating the general case for the capability approach (including, *inter alia*, the Aristotelian theory) from the special case for taking on *exclusively* this particular Aristotelian theory.

In fact, no matter whether we go the full Aristotelian way, which will also need a great deal of extension as a theory for practical evaluation, or take some other particular route, there is little doubt that the kind of *general* argument that Aristotle uses to motivate his approach does have a wider relevance than the defence of the particular form he gives to the nature of human good. This applies *inter alia* to Aristotle's rejection of opulence as a criterion of achievement (rejecting wealth and income as the standards), his analysis of *eudaimonia* in terms of valued activities (rather than relying on readings of mental states, as in some utilitarian procedures), and his assertion of the need to examine the processes through which human activities are chosen (thereby pointing towards the importance of freedom as a part of living).

11. Incompleteness and Substance

The Aristotelian critique points towards a more general issue, namely, that of the 'incompleteness' of the capability approach – both in generating substantive judgements and in providing a comprehensive theory of valuation. Quite different specific theories of value may be consistent with the capability approach, and share the common feature of selecting value-objects from functionings and capabilities. Further, the capability approach can be used with different methods of determining relative weights and different mechanisms for actual evaluation. The approach, if seen as a theory of algorithmic evaluation, would be clearly incomplete.[40]

It may well be asked: why pause at outlining a general approach, with various bits to be filled in, rather than 'completing the task'? The motivation underlying the pause relates to the recognition that an agreement on the usability of the capability approach – an agreement on the nature of the 'space' of value-objects – need not *presuppose* an agreement on how the valuational exercise may be completed. It is possible to disagree both on the exact *grounds* underlying the determination of relative weights, and on the *actual* relative weights chosen,[41] even when there is reasoned agreement on the general nature of the value-objects (in this case, personal functionings and capabilities). If reasoned agreement is seen as an important foundational quality central to political and social ethics,[42] then the case for the pause is not so hard to understand. The fact that the capability approach is consistent and combinable with several different substantive theories need not be a source of embarrassment.

Interestingly enough, despite this incompleteness, the capability approach does have considerable 'cutting power'. In fact, the more challenging part of the claim in favour of the capability approach lies in what it denies. It differs from the standard utility-based approaches in not insisting that we must value *only happiness* (and sees, instead, the state of being happy as one among several objects of value), or *only desire fulfilment* (and takes, instead, desire as useful but imperfect evidence – frequently distorted – of what the person herself values).[43] It differs also from other – non-utilitarian – approaches in not placing among value-objects *primary goods as such* (accepting these Rawlsian-focus variables only derivatively and instrumentally and only to the extent that these goods promote capabilities), or *resources as such* (valuing this Dworkinian perspective only in terms of the impact of resources on functionings and capabilities), and so forth.[44]

A general acceptance of the intrinsic relevance and centrality of the various functionings and capabilities that make up our lives does have substantial cutting power, but it need not be based on a prior agreement on the relative values of the different functionings or capabilities, or on a specific procedure for deciding on those relative values.

Indeed, it can be argued that it may be a mistake to move on relentlessly until one gets to exactly one mechanism for determining relative weights, or – to turn to a different aspect of the 'incompleteness' – until one arrives at exactly one interpretation of the metaphysics of value. There are substantive differences between different ethical theories at different levels, from the meta-ethical (involving such issues as objectivity) to the motivational, and it is not obvious that for substantive political and social philosophy it is sensible to insist that all these general issues be resolved *before* an agreement is reached on the choice of an evaluative space. Just as the utilization of actual weights in practical exercises may be based on the acceptance of a certain *range* of variability of weights (as I have tried to discuss in the context of the *use* of the capability approach[45]), even the general rationale for using such an approach may be consistent with some ranges of answers to foundational questions.

12. A Concluding Remark

In this paper I have tried to discuss the main features of the capability approach to evaluation: its claims, its uses, its rationale, its problems. I have also addressed some criticisms that have been made of the approach. I shall not try to summarize the main contentions of the paper, but before concluding, I would like to emphasize the plurality of purposes for which the capability approach can have relevance.

There are different evaluative problems, related to disparate value-purposes. Among the distinctions that are important is that between well-being and agency, and that between achievement and freedom. The four categories of intrapersonal assessment and interpersonal comparison that follow from these two distinctions (namely, well-being achievement, well-being freedom, agency achievement, and agency freedom) are related to each other, but are not identical. The capability approach can be used for each of these different types of evaluation, though not with equal reach. It is particularly relevant for the assessment of well-being – in the form of both achievement and freedom – and for the related problem of judging living standards.

As far as social judgements are concerned, the individual evaluations feed directly into social assessment. Even though the original motivation for using the capability approach was provided by an examination of the question 'equality of what?' (Sen, 1980), the use of the approach, if successful for equality, need not be confined to equality only.[46] The usability of the approach in egalitarian calculus depends on the plausibility of seeing individual advantages in terms of capabilities, and if that plausibility is accepted, then the same general perspective can be seen to be relevant for other types of social evaluation and aggregation.

The potentially wide relevance of the capability perspective should not come as a surprise, since the capability approach is concerned with showing the cogency of a particular *space* for the evaluation of individual opportunities and successes. In any social calculus in which individual advantages are constitutively important, that space is of potential significance.

Notes

1. This was in a Tanner Lecture given at Stanford University in May 1979 ('Equality of What?'), later published as Sen (1980). The case for focusing on capability was introduced here in the specific context of evaluating inequality. I have tried to explore the possibility of using the capability perspective for analysing other social issues, such as well-being and poverty (Sen, 1982*a*, 1983*c*, 1985*b*), liberty and freedom (Sen, 1983*a*, 1988*a*, 1992), living standards and development (Sen, 1983*b*, 1984, 1987*b*, 1988*b*), gender bias and sexual divisions (Kynch and Sen, 1983; Sen, 1985*c*, 1990*b*), and justice and social ethics (Sen, 1982*b*, 1985*a*, 1990*a*).
2. Though at the time of proposing the approach, I did not manage to seize its Aristotelian connections, it is interesting to note that the Greek word *dunamin*, used by Aristotle to discuss an aspect of the human good, which is sometimes translated as 'potentiality', can be translated also as 'capability of existing or acting' (see Liddell and Scott, 1977: 452). The Aristotelian perspective and its connections with the recent attempts at constructing a capability-focused approach have been illuminatingly discussed by Martha Nussbaum (1988).
3. See the contributions of Roemer (1982, 1986), Streeten (1984), Beitz (1986), Dasgupta (1986, 1988, 1989), Hamlin (1986), Helm (1986), Zamagni (1986), Basu (1987), Brannen and Wilson (1987), Hawthorn (1987); Kanbur (1987), Kumar (1987), Muellbauer (1987), Ringen (1987), B. Williams (1987), Wilson (1987), Nussbaum (1988, 1990), Griffin and Knight (1989*a*, 1989*b*), Riley (1988), Cohen (1990), and Steiner (1990). On related matters, including application, critique, and comparison, see also de Beus (1986), Kakwani (1986), Luker (1986), Sugden (1986), Asahi (1987), Delbono (1987), Koohi-Kamali (1987), A. Williams (1987), Broome (1988), Gaertner (1988), Stewart (1988), Suzumura (1988), de Vos and Hagennars (1988), Goodin (1985, 1988), Hamlin and Pettit (1989), Seabright (1989), Hossain (1990), and Schokkaert and van Ootegem (1990), among others.
4. If there are *n* relevant functionings, then a person's extent of achievement of all of them respectively can be represented by an *n*-tuple. There are several technical problems in the representation and analysis of functioning *n*-tuples and capability sets, on which see Sen (1985*b*: chs. 2, 4, and 7).

5. Bernard Williams (1987) raises this issue in his comments on my Tanner Lectures on the standard of living (pp. 98–101); on which see also Sen (1987*b*: 108–9). On the inescapable need for evaluation of different functioning and capabilities, see Sen (1985*b*: chs. 5–7). Just as the concentration on the commodity space in real-income analysis does not imply that every commodity must be taken to be equally valuable (or indeed valuable at all), similarly focusing on the space of functioning does not entail that each functioning must be taken to be equally valuable (or indeed valuable at all).
6. On this and on other formulations and uses of dominance ranking, see Sen (1970: chs. 1*, 7*, 9*).
7. On the crucial role of the informational basis, and on the formulation and use of informational constraints, see Sen (1970, 1977) and d'Aspremont and Gevers (1977).
8. Welfarism requires that a state of affairs must be judged by the individual utilities in that state. It is one of the basic components of utilitarianism (the others being 'sum-ranking' and 'consequentialism'); on the factorization, see Sen (1982*a*) and Sen and Williams (1982).
9. Being happy and getting what one desires may be *inter alia* valued in the capability approach, but unlike in utilitarian traditions, they are not seen as the measure of all values.
10. On this, see Sen (1970, 1982*a*, 1987*a*). In many contexts, the mathematical representations should take the form of 'partial orderings' or 'fuzzy' relations. This is not, of course, a special problem with the capability approach, but applies generally to conceptual frameworks in social, economic, and political theory.
11. The belief in this possibility seems to play a part in Robert Sugden's (1986) criticism of what he sees as my approach to capability evaluation, namely, a 'general strategy of trying to derive the value of a *set* of functioning vectors from prior ranking of the vectors themselves' (p. 821). He argues in favour of judging 'the value of being free to choose from a range of possible lives' *before* taking 'a view on what constitutes a valuable life'. This criticism is, in fact, based on a misunderstanding of the approach proposed, since it has been a part of my claim (on which more presently) that the judgement of the quality of life and the assessment of freedom have to be done *simultaneously* in an integrated way, and, in particular, that 'the quality of life a person enjoys is not merely a matter of what he or she achieved, but also of what options the person has had the opportunity to choose from' (Sen, 1985*b*: 69–70). But the point at issue in the present context is the possibility of judging a *range* of choice independently of the value characteristics of the *elements* in that range. It is this possibility that I am disputing.
12. Even this can be questioned when an expanded menu causes confusion, or the necessity to choose between a larger set of alternatives is a nuisance. But such problems can be dealt with through *appropriate* characterization of all the choices one has or does not have. This must include the consideration of the overall choice of having *or* not having to choose among a whole lot of relatively trivial alternatives (e.g. the choice of telling the telephone company to shut out mechanically dialled calls from sales agents offering a plethora of purchasing

options). The issues involved in this kind of complex evaluation, incorporating choices over choices, are discussed in Sen (1992).

13. For an illuminating axiomatic derivation of the number-counting method of freedom evaluation, see Pattanaik and Xu (1990).
14. The unacceptability of this kind of number-counting evaluation of freedom is discussed in Sen (1985*b*). For an assessment of the axiomatic foundations of this and other methods of evaluation of freedom, see Sen (1991).
15. This type of case also shows why the set-inclusion ranking is best seen as a 'weak' relation of 'no worse than' or 'at least as good as', rather than as the 'strict' relation of 'better than'. Adding the option of 'slicing one's toes' to the set of valued options a person already has may not *reduce* her freedom (since one can reject toe-slicing), but it is hard to take it to be a strict *increase* in that person's freedom.
16. As was argued earlier, the relation is two-sided, and the evaluation of the freedom to lead a life and the assessment of the life led (including choosing freely) have to be done simultaneously, in a desegregated way.
17. Since a person's agency objectives will typically include, *inter alia*, his or her own well-being, the two will to some extent go together (e.g. an increase in well-being, other things being equal, will involve a higher agency achievement). In addition, a failure to achieve one's *non*-well-being objectives may also cause frustration, thereby reducing one's well-being. These and other connections exist between well-being and agency, but they do not make the two concepts congruent – nor isomorphic in the sense of generating the same orderings. Similarly, more freedom (either to have well-being or to achieve one's agency goals) may lead one to end up achieving more (respectively, of well-being or of agency success), but it is also possible for freedom to go up while achievement goes down, and vice versa. We have here four *interdependent* but *non-identical* concepts. These distinctions and their inter-relations are discussed more fully in Sen (1985*a*, 1992).
18. See Adam Smith (1776: Vol. ii, Bk V, ch. 2 (section on 'Taxes upon Consumable Commodities')), in Campbell and Skinner (1976), 469–71.
19. Rock Edicts XIII at Erragudi, statement VII. For a translation and discussion, see Sircar (1979: 34).
20. This view may be disputed by considering a different way of drawing the line between well-being and living standards. One common approach is to relate the assessment of living standards only to real incomes and to 'economic' or 'material' causes. On this see A. C. Pigou (1920); and on the conceptual differences see Bernard Williams (1987). But the Pigovian view has problems of its own. For example, if one has a disability that makes one get very little out of material income or wealth, or if one's life is shattered by an inconvenient and incurable illness (e.g. kidney problems requiring extensive dialysis), it is hard to claim that one's standard of living is high just because one is well heeled. I have discussed this question and related matters in Sen (1987*b*: 26–9, 109–10).
21. For formal characterizations, see Sen (1985*b*: chs. 2 and 7).
22. On this see Sen (1985*b*: 60–1). The distinguished element can be the *achieved* one (as in this case), or more specifically the *chosen* one (if there is a choice exercise in determining what happens), or the *maximal* one (in terms of some

criterion of goodness). The three will coincide if what is achieved is achieved through choice, and what is chosen is chosen through maximization according to that criterion of goodness.

23. As was argued earlier in dealing with responsible adults, it may be appropriate to see the claims of individuals on society in terms of the *freedom* to achieve well-being (and thus in terms of real opportunities) rather than in terms of *actual achievements*. If the social arrangements are such that a responsible adult is given no less freedom (in terms of set comparisons) than others, but he still 'muffs' the opportunities and ends up worse off than others, it is possible to argue that no particular injustice is involved. On this and related matters, see Sen (1985*a*).
24. The same capability set can then be used for the evaluation of both 'well-being achievement' (through *elementary evaluation*, concentrating on the achieved element) and 'well-being freedom' (through *non-elementary set evaluation*).
25. Thus, in standard consumer theory, set evaluation takes the form of elementary evaluation. For particular departures from that tradition, see Koopmans (1964) and Kreps (1979). In the Koopmans-Kreps approach, however, the motivation is not so much to see living freely as a thing of intrinsic importance, but to take note of uncertainty regarding one's own future preference by valuing – instrumentally – the advantage of having more options in the future. On the motivational contrasts, see Sen (1985*a*, 1985*b*).
26. As was argued in an earlier paper, 'the "good life" is partly a life of genuine choice, and not one in which the person is forced into a particular life – however rich it might be in other respects' (Sen, 1985*b*: 69–70).
27. See Sen (1985*b*, 1988*a*, 1991), Suppes (1987), Pattanaik and Xu (1990).
28. The characteristics and relevance of 'refined functioning' have been discussed in Sen (1985*a*, 1988*a*).
29. These issues are discussed in Sen (1985*a*, 1987*b*).
30. The 'basic needs' literature is extensive. For a helpful introduction, see Streeten *et al.* (1981). In a substantial part of the literature, there is a tendency to define basic needs in the form of needs for *commodities* (e.g. for food, shelter, clothing, health care), and this may distract attention from the fact that these commodities are no more than the *means* to real ends (inputs for valuable functionings and capabilities). On this question, see Streeten (1984). The distinction is particularly important since the relationship between commodities and capabilities may vary greatly between individuals even in the same society (and of course between different societies). For example, even for the elementary functioning of being well-nourished, the relation between food intake and nutritional achievements varies greatly with metabolic rates, body size, gender, pregnancy, age, climatic conditions, epidemiological characteristics, and other factors (on these and related matters see Drèze and Sen, 1989). The capability approach can accommodate the real issues underlying, the concern for basic needs, avoiding the pitfall of 'commodity fetishism'.
31. The importance of socially recognized ideas of 'urgency' has been illuminatingly discussed by Thomas Scanlon (1975).
32. On this see Sen (1980). To avoid confusion, it should also be noted that the term 'basic capabilities' is sometimes used in quite a different sense from the one

specified above, e.g. as a person's *potential* capabilities that *could* be developed, whether or not they are actually realized (this is the sense in which the term is used, for example, by Martha Nussbaum (1988)).

33. While the notion of basic capabilities was used in Sen (1980, 1983*c*), in later papers the capability approach has been used without identifying certain capabilities as 'basic' and others as not so (see e.g. Sen, 1984, 1985*a*, 1985*b*). This point is relevant to G. A. Cohen's distinction between focusing on what he calls 'midfare' and on functioning and capabilities. There are more important distinctions to explore (to be taken up in Section 9), but the contrasts look artificially sharper if the capability approach is seen as being confined *only* to the analysis of basic capabilities.
34. On this see Sen (1983*c*). See also Drèze and Sen (1989) and Hossain (1990).
35. Technically, what is being used in this analysis is the 'inverse function', taking us back from specified capability levels to necessary incomes, given the other influences on capability. This procedure will not be usable, in this form, if there are people who are so handicapped in terms of personal characteristics that no level of income will get them to reach minimally acceptable basic capabilities; such people would then be invariably identified as poor.
36. See, particularly, Smith (1776) and Marx (1844). The connections are discussed in Sen (1984, 1985*a*, 1987*b*).
37. See particularly *The Nicomachean Ethics*, Bk I, s. 7; in the translation by Ross (1980: 12–14).
38. On this see Nussbaum (1990).
39. Nussbaum (1988: 176).
40. This relates to one part of the critique presented by Beitz (1986).
41. On this see Sen (1985*b*: chs. 5–7).
42. On this question, see Rawls (1971), Scanlon (1982), B. Williams (1985).
43. For comparisons and contrasts between the capability approach and utilitarian views, see Sen (1984, 1985*a*).
44. See Rawls (1971, 1988*a*, 1988*b*), Dworkin (1981), and Sen (1980, 1984, 1990*a*).
45. See Sen (1985*b*); on the general strategy of using 'intersection partial orders', see Sen (1970, 1977).
46. Corresponding to 'equality of what?', there is, in fact, also the question: 'efficiency of what?'

References

Aristotle (4th c. BC). *The Nicomachean Ethics*: see Ross (1980).

Arneson, R. (1987). 'Equality and Equality of Opportunity for Welfare', *Philosophical Studies*, S4.

Asahi, J. (1987). 'On Professor Sen's Capability Theory', mimeographed, Tokyo.

Basu, K. (1987). 'Achievements, Capabilities and the Concept of Well-being', *Social Choice and Welfare*, 4.

Beitz, C. R. (1986). 'Amartya Sen's *Resources, Values and Development*', *Economics and Philosophy*, 2.

Brannen, J., and Wilson, G. (eds.) (1987). *Give and Take in Families*. London: Allen and Unwin.

Broome, J. (1988). Review of *On Ethics and Economics* and *The Standard of Living in London Review of Books*.
Campbell, R. H., and Skinner, A. S. (eds.) (1976). Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.
Cohen, G. A. (1989). 'On the Currency of Egalitarian Justice', *Ethics*, 99.
—— (1990). 'Equality of What? On Welfare, Goods and Capabilities', *Recherches Economiques de Louvain*, 56.
Culyer, A. J. (1985). 'The Scope and Limits of Health Economies', *Okonomie des Gesundheitswesens*.
Dasgupta, P. (1986). 'Positive Freedom, Markets and the Welfare State', *Oxford Review of Economic Policy*, 2.
—— (1988). 'Lives and Well-being', *Social Choice and Welfare*, 5.
—— (1989). 'Power and Control in the Good Polity', in Hamlin and Pettit (1989).
d'Aspremont, C, and Gevers, L. (1977). 'Equity and the Informational Basis of Collective Choice', *Review of Economic Studies*, 46.
deBeus, Jos (1986). 'Sen's Theory of Liberty and Institutional Vacuum', University of Amsterdam.
deVos, K., and Hagenaars, Aldi J. M. (1988). 'A Comparison between the Poverty Concepts of Sen and Townsend', mimeographed, University of Leiden.
Deaton, A., and Muellbauer, J. (1980). *Economics and Consumer Behaviour*. Cambridge: Cambridge University Press.
Delbono, F. (1987). Review article on *Commodities and Capabilities, Economic Notes*.
Drèze, J., and Sen, A. (1989). *Hunger and Public Action*. Oxford: Clarendon Press.
Dworkin, R. (1981). 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs*, 10.
Elster, J., and Hylland, A. (eds.) (1986). *Foundations of Social Choice Theory*. Cambridge: Cambridge University Press.
Gaertner, W. (1988). Review of *Commodities and Capabilities in Zeitschriften fur National Okommnomia*, 48.
Goodin, R. E. (1985). *Political Theory and Public Policy*. Chicago and London: University of Chicago Press.
—— (1988). *Reasons for Welfare: Political Theory of the Welfare State*. Princeton: Princeton University Press.
Griffin, J. (1986). *Well-being*. Oxford: Clarendon Press.
Griffin, K., and Knight, J. (eds.) (1989*a*). 'Human Development in the 1980s and Beyond', special number, *Journal of Development Planning*, 19.
—— (1989*b*). 'Human Development: The Case for Renewed Emphasis', *Journal of Development Planning*, 19.
Hamlin, A. P. (1986). *Ethics, Economics and the State*. Brighton: Wheatsheaf Books.
Hamlin, A., and Pettit, P. (eds.) (1989). *The Good Polity: Normative Analysis of the State*. Oxford: Blackwell.
Hare, R. (1981). *Moral Thinking*. Oxford: Clarendon Press.
Harrison, R. (ed.) (1979). *Rational Action*. Cambridge: Cambridge University Press.
Hawthorn, G. (1987). 'Introduction' in Sen (1987b).
Helm, D. (1986). The Assessment: The Economic Border of the State', *Oxford Review of Economic Policy*, 2.
Hossain, I. (1990). *Poverty as Capability Failure*. Helsinki: Swedish School of Economics and Business Administration.

Kakwani, N. (1986). *Analysing Redistribution Policies*. Cambridge: Cambridge University Press.
Kanbur, R. (1987). 'The Standard of Living: Uncertainty, Inequality and Opportunity', in Sen (1987*b*).
Koohi-Kamali, F. (1988), 'The Pattern of Female Mortality in Iran and some of its Causes', Applied Economics Discussion Paper 62, Oxford Institute of Economics and Statistics.
Koopmans, T. C. (1964). 'On Flexibility of Future Preference', in M. W. Shelly and G. L. Bryan (eds.), *Human Judgments and Optimality*. New York: Wiley.
Kreps, D. (1979). 'A Representation Theorem for Preference for Flexibility', *Econometrica*, 47.
Kumar, B. G. (1987). Poverty and Public Policy: Government Intervention and Level in Kerala, India. D.Phil, dissertation, Oxford University.
Kynch, J., and Sen, A. (1983). 'Indian Women: Well-being and Survival', *Cambridge Journal of Economics*, 7.
Llddell, H. G., and Scott, R. (1977). *A Greek-English Lexicon*, extended by H. S. Jones and R. McKenzie. Oxford: Clarendon Press.
Luker, W. (1986). 'Welfare Economics, Positivist? Idealism and Quasi-Experimental Methodology', mimeographed, University of Texas, Austin.
Marx, K. (1844). *Economic and Philosophic Manuscripts*. English translation. London: Lawrence and Wishart, 1977.
McMurrin, S. M. (ed.) (1980). *Tanner Lectures on Human Values*, i. Salt Lake City: University of Utah Press, and Cambridge: Cambridge University Press.
Muellbauer, J. (1987). 'Professor Sen on the Standard of Living', in Sen (1987*b*).
Nussbaum, M. (1988). 'Nature, Function, and Capability: Aristotle on Political Distribution', *Oxford Studies in Ancient Philosophy*, suppl. vol.
——— (1990). 'Non-Relative Virtues: An Aristotelian Approach', *Midwest Studies in Philosophy*, 13; revised version in this volume.
Pattanaik, P. K., and Xu, Yongsheng (1990). 'On Ranking Opportunity Sets in Terms of Freedom of Choice', *Recherches Economiques de Louvain*, S6.
Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.
Rawls, J. (1971). *A Theory of Justice*. Cambridge, Mass.: Harvard University Press, and Oxford: Clarendon Press.
——— (1988*a*). 'Priority of Right and Ideas of the Good', *Philosophy and Public Affairs*, 17.
——— (1988*b*). 'Reply to Sen', mimeographed, Harvard University.
——— *et al.* (1987). *Liberty, Equality, and Law: Selected Tanner Lectures on Moral Philosophy*, ed. S. McMurrin. Cambridge: Cambridge University Press, and Salt Lake City: University of Utah Press.
Riley, J. (1988). *Liberal Utilitarianism: Social Choice Theory and J. S. Mill's Philosophy*. Cambridge: Cambridge University Press.
Ringen, J. (1987). *The Possibility of Politics: A Study of the Economy of the Welfare State*. Oxford: Clarendon Press.
Roemer, J. (1982). *A General Theory of Exploitation and Class*. Cambridge, Mass.: Harvard University Press.
——— (1986). 'An Historical Materialist Alternative to Welfarism', in Elster and Hylland (1986).
Ross, D. (1980). Aristotle, *The Nicomachean Ethics*. The World's Classics. Oxford: Oxford University Press.

Scanlon, T. M. (1979). 'Preference and Urgency', *Journal of Philosophy*, 72.
——— (1982). 'Contractualism and Utilitarianism', in Sen and Williams (1982).
Schokkaert, E., and van Ootegem, L. (1990). 'Sen's Concept of the Living Standard Applied to the Belgian Unemployed', *Recherches Economiques de Louvain*, S6.
Seabright, P. (1989). 'Social Choice and Social Theories', *Philosophy and Public Affairs*, 18.
Sen, A. K. (1970). *Collective Choice and Social Welfare*. San Francisco: Holden-Day. Republished Amsterdam: North-Holland, 1979.
——— (1977). 'On Weights and Measures: Informational Constraints in Social Welfare Analysis', *Econometrica*, 45.
——— (1980). 'Equality of What?' (1979 Tanner Lecture at Stanford), in McMurrin (1980); repr. in Sen (1982a) and Rawls *et al.* (1987).
——— (1982*a*). *Choice, Welfare and Measurement*. Oxford: Blackwell, and Cambridge, Mass.: MIT Press.
——— (1982*b*). 'Rights and Agency', *Philosophy and Public Affairs*, 11.
——— (1983*a*). 'Liberty and Social Choice', *Journal of Philosophy*, 80.
——— (1983*b*). 'Development: Which Way Now?', *Economic Journal*, 93; repr. in Sen (1984).
——— (1983*c*). 'Poor, Relatively Speaking, *Oxford Economic Papers*, 35; repr. in Sen (1984).
——— (1984). *Resources, Values and Development*. Oxford: Blackwell, and Cambridge, Mass.: Harvard University Press.
——— (1985*a*). 'Well-being, Agency and Freedom: The Dewey Lectures 1984', *Journal of Philosophy*, 82.
——— (1985*b*). *Commodities and Capabilities*. Amsterdam: North-Holland.
——— (1985*c*). 'Women, Technology and Sexual Divisions', *Trade and Development*, 6.
——— (1987*a*). *On Ethics and Economics*. Oxford: Blackwell.
——— (1987*b*). *The Standard of Living* (1985 Tanner Lectures at Cambridge, with contributions by Keith Hart, Ravi Kanbur, John Muellbauer, and Bernard Williams, edited by G. Hawthorn). Cambridge: Cambridge University Press.
——— (1988*a*). 'Freedom of Choice: Concept and Content', *European Economic Review* 32.
——— (1988*b*). 'The Concept of Development', in H. Chenery and T. N. Srinivasan (eds.), *Handbook of Development Economics*. Amsterdam: North-Holland.
——— (1990*a*). 'Justice: Means versus Freedoms', *Philosophy and Public Affairs*, 19.
——— (1990b). 'Gender and Cooperative Conflicts', in Tinker (1990).
——— (1991). 'Preference, Freedom and Social Welfare', *Journal of Econometrics*, 50.
——— (1992). *Inequality Reexamined*. Oxford: Clarendon Press.
——— and Williams, B. (eds.) (1982). *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
Sircar, D. C. (1979). *Asokan Studies*. Calcutta: Indian Museum.
Smith, Adam (1776). *An Inquiry into the Nature of Causes of the Wealth of Nations*: see Campbell and Skinner (1976).
Steiner, H. (1986). 'Putting Rights in their Place: An Appraisal of Amartya Sen's Work on Rights', mimeographed, University of Manchester.
Stewart, F. (1988). 'Basic Needs Strategies, Human Rights and the Right to Development', mimeographed, Queen Elizabeth House.

Streeten, P. (1984). 'Basic Needs: Some Unsettled Questions', *World Development*, 12.

Streeten, P., *et al.* (1981). *First Things First; Meeting Basic Needs in Developing Countries.* New York: Oxford University Press.

Sugden, R. (1986). Review of *Commodities and Capabilities, Economic Journal*, 96.

Suppes, P. (1987). 'Maximizing Freedom of Decision: An Axiomatic Analysis', in G. R. Feiwel (ed.), *Arrow and the Foundations of Economic Policy*. New York: New York University Press.

Suzumura, K. (1988). Introduction to the Japanese translation of *Commodities and Capabilities.* Tokyo: Iwanami.

Tinker, I. (ed.) (1990). *Persistent Inequalities.* New York: Oxford University Press.

Williams, A. (1985). 'Economics of Coronary Bypass Grafting', *British Medical Journal*, 291.

——— (1991). 'What is Health and Who Creates it?', in J. Hutton *et al.* (eds.), *Dependency to Enterprise.* London: Routledge.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. London: Fontana, and Cambridge, MA: Harvard University Press.

——— (1987). The Standard of Living: Interests and Capabilities', in Sen (1987*b*).

Wilson, G. (1987). *Money in the Family*. Aldershot: Avebury.

Zamagni, S. (1986). 'Introduzione', in A. Sen, *Scelta, Benessere, Equita.* Bologna: Il Mulino.

PART FOUR

BRANCHES AND SCHOOLS OF ECONOMICS AND THEIR METHODOLOGICAL PROBLEMS

Economics is a diverse undertaking consisting of different branches and schools. Within what one might call "mainstream economics" – which is increasingly dominant – one finds many different activities, each with its own methodological peculiarities and problems. Discussions of many intriguing branches of mainstream economics could have been included in Part IV. Inquiries such as behavioral economics, industrial management, the economics of information, labor economics, game theory, and so forth all raise distinctive and important methodological questions.

Because of space constraints, Part IV contains methodological discussions of only four branches of mainstream economics. In Chapter 16, Kevin D. Hoover examines some of the philosophical issues concerning contemporary econometrics, including some of the problems involved in making causal inferences. In Chapter 17, Hoover addresses the question of how closely linked macroeconomics should be to microeconomics, which has long been one of the central methodological questions concerning macroeconomics. In Chapter 18, Vernon Smith, who won a Nobel Prize mainly for his work in experimental economics probes some of the central methodological questions raised by this relatively new and exciting field, whereas in Chapter 19, Colin F. Camerer explores the possibility that inquiries into neurology might guide economic theorizing.

The remaining two chapters in Part IV discuss two of the most important approaches to economics that compete with mainstream economics. In Chapter 20, James M. Buchanan and Viktor J. Vanberg provide an introduction to the distinctive subjective methodology of Austrian economics and, in Chapter 21, Geoffrey M. Hodgson provides an overview of the way in which institutionalist or so-called evolutionary economists approach the subject.

SIXTEEN

Econometrics as Observation

The Lucas Critique and the Nature of Econometric Inference

Kevin D. Hoover

Kevin Hoover (1955–) received a D.Phil. in economics from Oxford University after an undergraduate major in philosophy, and his work reflects this dual competence. He has contributed both to contemporary economics (especially macroeconomics) and to economic methodology, serving for a decade as the editor of *The Journal of Economic Methodology*. After more than two decades at the University of California, Davis, Hoover is now a professor of economics and a professor of philosophy at Duke University.

1. The Lucas Critique

Perhaps the principal challenge to the use of econometric models in economic analysis is the policy non-invariance argument, popularly known as the 'Lucas critique'. Robert Lucas (1976) attacks the use of econometric models as bases for the evaluation of policy on the grounds that the estimated equations of such models are unlikely to remain invariant to the very changes in policy that the economist seeks to evaluate. The argument is originally cast as an implication of rational expectations. Among the constraints people face are the policy rules of the government. If people are rational, then, when these rules change, and if the change is correctly perceived, they take proper account of the change in adjusting their behavior. The rational expectations hypothesis implies that changes in policy will in fact be correctly perceived up to a serially uncorrelated error.

The Lucas critique challenges macroeconometrics along two related paths. First, it suggests that existing models are useless for evaluating prospective changes in policy; second, it suggests that existing models are not accurate representations of even the current structure of the economy.

Journal of Economic Methodology, vol. 1 (1994): 65–80. Reprinted by permission of Taylor and Francis Ltd.

I want to suggest that the first path can be thought of as a denying that econometric models typically capture the true causal structure of the economy, and that the second path can be thought of as denying that the models are identified in the econometrician's usual sense. These are closely related ideas, but they are not identical. Christopher Sims (1980: 1) faults large-scale econometric models for relying on 'incredible' identifying restrictions; yet Sims (1982; cf. Hoover 1988a: 197–202) asserts their usefulness, in a highly restricted sense, in evaluating alternative policies.

The concept of cause and which concepts are appropriate in which circumstances are hotly debated among some econometricians and economic methodologists (see, e.g., Granger 1980; Leamer 1985; Zellner 1979, and the supplement to the *Journal of Econometrics* 1988). The notion of cause as control, however, seems naturally appropriate in the debate over the Lucas critique. Roughly, a change in policy causes a change in the economy if the change in policy can be used to control that aspect of the economy. I have argued elsewhere that a good rendering of this notion of causality is to be found in J. L. Mackie's (1980) conditional analysis of causality.[1]

Mackie defines a cause to be an Insufficient, Non-redundant member of a set of Unnecessary but Sufficient conditions for the effect. This is often called the INUS condition. Simply put it says that *a cause* is a critical part of one of the possibly numerous alternative combinations of circumstances that imply an effect. The details of Mackie's analysis are not important in the current discussion. What is important is to notice that causal relations are captured in this analysis by (contrary-to-fact) conditional propositions; that is, by statements of the form, 'if it were true that the economy was at full employment and the money supply increased by 10 percent, then it would be true that prices would rise by 10 percent'. Such a proposition sustains our belief that increases in the money supply cause increases in the price level. Yet its correctness is not challenged by our failure to observe the antecedents to be fulfilled.[2]

The conditional analysis of causality is related to both invariance and control. Consider a putatively correct conditional proposition whose antecedents are not fulfilled. If it happens that when the antecedents are in fact fulfilled, the conditional proposition is no longer correct, it was wrong to assert it in the first place. Any conditional proposition asserts the existence of a disposition. The properties of such a disposition cannot depend on whether or not it is in fact actualized. Thus, 'a diamond is hard enough to scratch glass' asserts a disposition. It is correct whether or not a diamond is ever used to scratch glass. But it would not be correct if hardness were attributed to the diamond only when no one attempted to scratch glass. Such dispositions necessarily presuppose invariance. And the conditional analysis of causality is partly the assertion that causal relations are invariant to attempts to use them to control the effects. Policy interventions are then connected to effects in the economy through invariant causal relations. The Lucas critique can thus be read as the claim that existing macroeconometric models do not isolate causal relations – i.e., they do not assert correct conditional propositions.

An INUS condition is not the complete cause of its effect. In general, we are interested in some INUS conditions but wish not to pay direct attention to others. Both Federal Reserve policy and the institutional structure of Wall Street may be INUS conditions of the term structure of interest rates; but a bond trader is directly interested in Fed policy and generally relegates institutional structure to what Mackie calls the 'causal field'. The causal field is simply the set of INUS conditions, which either do not change or which serve as boundary conditions for the problem at hand. To say that a causal relation is invariant to interventions of control leaves unstated the caveat, 'within a particular causal field'.

In a related paper (Hoover 1990), I demonstrate that one can link Mackie's conditional analysis of causality with a characterization of causality in systems of equations first developed by Herbert Simon (1953). Causality in a linear system of equations can be associated in Simon's analysis with block recursion between variables.[3] A variable ordered ahead of another in the recursion is said to cause the other. Since each equation can be thought of as a conditional proposition with its parameters and variables as antecedents, it is easy to surmise that some sort of mapping exists between Mackie's analysis and Simon's. Simon demonstrates that when a system of equations is causally ordered it is identified econometrically. The Lucas critique can therefore also be seen as the claim that existing macroeconometric models are in fact not identified.

2. Identification and Causality

Models are formal representations that are meant to capture or mimic reality. Causality can be defined within the context of the model as Simon does or as a property of the world as Mackie does. The distinction between causal relations as they truly are and representations of them suggests that an important empirical problem is how we might infer from data what the underlying causal relations are; that is, how we might learn how our formal representations should be constructed or, given a tentative model, how we might check whether it has been constructed satisfactorily. The invariance property of causal relations might be exploited as a source of information. If models prove to be invariant to a wide class of interventions, such as changes in policy or in institutional structure, we have some grounds for tentatively accepting them as representations of the true causal structure – at least within the limits of the particular causal field justified by the range of the interventions. Equally, even if we do not postulate a causal model, we may by observing the behavior of statistical relations between observable variables under known interventions still be able to discern some of the causal links that any satisfactory model must represent.

These two approaches to learning about causal relations by means of the invariance property – that is, testing models and seeking restrictive criteria for any satisfactory model – are quite different, although not necessarily competitive. Both are strategies to secure invariance in economic models and, thus, to lend support to the attribution of verisimilitude to them. In section 3 below we will examine both approaches under the headings 'apriorism' and 'econometrics as observation'.

The Lucas critique and the new classical worldview generally suggest that a strong argument against the observational approach to assessing causal relations is that observed regularities are not autonomous but merely derivative. New classical economists argue that, since some of the important variables in the true causal ordering such as expectational variables, are intrinsically unobservable, observable statistical relations cannot be autonomous. Furthermore, even the true causal relations will be complicated, especially if, as is an article of faith among the new classicals, people use all the available information economically. Thus, they argue that it is the a priori approach which has the best hope of success, because it is only if one knows how to specify the linkages among observable variables, including those implied by their statistical relations to unobservable variables, that invariance can ever be observed.

To the econometrician, perhaps the most important obstacle to estimating models of economic processes is the so-called 'identification problem'. The root of the problem is this: economic theory uses variables to describe economic processes which are not observable; observable variables are the outcome of interactions among these unobservables; and without further information it is, in general, not possible to infer the behavior of the unobservables from the observables. The paradigm identification problem is the simple system in which desired or planned supply is an increasing function of price alone and desired or planned demand is a decreasing function of price alone. Together these functions determine an equilibrium observable outcome – the amount sold. The identification problem is that from observing this single price/quantity combination, one cannot infer what the underlying functions are. The problem is no easier in the presence of random shocks to the functions; the single observed point is then simply replaced by a scatter of points randomly distributed about the equilibrium. If we have enough additional information, say that the variance of the random shocks to the supply curve is much greater than that to the demand curve or that supply is also a function of, say, rainfall, then it may be possible to infer the shape of the underlying functions (subject to some random error), in this case because the movements of the supply curve from random shocks or variability in rainfall force the observed price/quantity combinations to trace out the demand curve.

The Lucas critique is a variation on the theme of the identification problem. Just as we find observed price/quantity patterns (a point, a scatter or a line, depending on the nature of the actual underlying relation) jumping about when underlying but unaccounted for factors, such as the level of rainfall, change, we notice that estimated (presumed) behavioral functions appear not to be stable in the face of policy changes not accounted for in the estimate. When seen in this light, the Lucas critique clearly did not originate with Lucas and deserves to bear his name only because he brought the invariance problem home to most economists more forcibly than any earlier author and because it serves as a convenient shorthand.

Lucas (1976) himself claims no originality for the non-invariance argument, suggesting that it is implicit in the work of Frank Knight, Milton Friedman and John Muth. Lucas does not, however, notice the explicit statement in Trygve Haavelmo's famous paper, 'The probability approach in econometrics' (1944). Haavelmo compares the estimation of simple econometric relations to working out the relation between the amount of throttle and the speed of a car on a flat track under uniform conditions. The relation may be precise; but change the surrounding conditions (e.g., take the car off the track or allow the engine to get out of tune), and the relation will almost certainly break down. Haavelmo (1944: 28) contrasts the lack of autonomy of such empirical regularities with such things as the laws of thermodynamics, friction and so forth, which are autonomous because they 'describe the functioning of some parts of the mechanism *irrespective* of what happens in some *other* parts'. What Haavelmo suggests, in the terminology of Zellner (1979), is that non-autonomous relations are not lawlike; they do not represent the underlying causal ordering. It will not do to overstate the case; so Haavelmo goes on to argue that autonomy is a matter of degree. Again this can be rephrased: a causal ordering is claimed to be invariant only with respect to a particular causal field, which may itself be of a broad or a narrow compass. In some cases, then, it may be useful to see the non-invariance of a relation (in the car example, for instance) as a change in the causal field. Haavelmo does not himself state the invariance problem in terms of representing causal relations, but Simon (1953: 25–7) puts it in exactly such terms and is, therefore, another precursor of Lucas. Simon (1953: 27) writes:

> causal ordering is a property of models that is invariant with respect to interventions within the model, and structural equations are equations that correspond to specified possibilities of intervention.
>
> But it is precisely this same notion of intervention, and this same distinction between structural and nonstructural equations, that lies at the root of the identifiability concept.

Although Lucas was not the first to recognize the invariance problem explicitly, his own important contribution to it is to observe that one of the relations frequently omitted from putative causal representations is that of the formation of expectations. He notes, further, that the formation of expectations may depend upon people's understanding of the causal structure of the economy in general and of the process of policy formation in particular. This is why the rational expectations hypothesis is often linked with the Lucas critique. The important point about the analysis of non-invariance provided by Haavelmo and Simon is to remind us that rational expectations is simply one means by which causal relations may be linked; and, in general, it is the omission of *any* causal relation related to those remaining that produces non-invariance.[4] Lucas and his predecessors have, then, diagnosed a problem for economic analysis. We must now turn to proposed cures.

3. Strategies for Securing Invariance

Apriorism

The invariance problem can be treated in two complementary ways: as a problem of representing causal relations or as a problem of identifiability. In discussing the problem of securing identification, econometric textbooks recommend that restrictions be imposed a priori on the basis of economic theory.[5] Thus, the second approach explains why the most frequently proposed strategy for securing invariance in econometric models subject to the

Lucas critique is to structure such models around restrictions deduced a priori from economic theory. This *apriorism* is the received view of two generations of economists. Apriorism comes in two forms – strong and weak. Strong apriorism can be traced back to Tjalling Koopmans' attack, in his paper, 'Measurement without theory' (1947), on the atheoretical methods of Wesley Mitchell and his colleagues at the National Bureau of Economic Research, The strong apriorist view has been forcibly restated by Thomas F. Cooley and Stephen F. LeRoy (1985). They argue that work by the Cowles Commission in the 1940s – particularly by Simon and Koopmans – established the need to impose identifying restrictions on econometric equations if estimates of structural parameters are to be obtained. Furthermore, they argue that such restrictions are untestable. The only basis one can possibly have for imposing them, therefore, is that they are derived from a well-articulated theory acceptable a priori. By theory Cooley and LeRoy seem to mean a well-specified, consistent optimization problem. They go on to observe that the rational expectations hypothesis, which often generates a high degree of interdependence between variables in theoretical models, suggests that even economic theory cannot provide the restrictions needed to identify large-scale macroeconometric models. In this they are at one with Lucas and Sargent and other new classical economists, who argue that, only if economic theories are grounded in well-specified optimization problems, taking tastes and technology alone as given, will they be secure from the invariance problem (Lucas and Sargent 1979; Lucas 1981: Introduction).

The econometric analogue of the apriorist's belief in the dominance of economic theory is the familiar view that the *objective* significance of an econometric result varies with the means by which that result is obtained. On this view, good econometrics starts with a hypothesis derived from theory which dictates certain expected signs and significance levels of coefficients. It then estimates a regression and checks whether the result accords with the a priori expectation – if yes, fine; if no, back to the theoretical drawing board.[6] Bad econometrics tries out arbitrary (i.e., atheoretical) specifications until the results suit the investigator's prior beliefs. The strong apriorist's view of econometrics has the implausible, counterintuitive result that, if I happen to estimate a confirming regression at the first go, my theory is supported; while, if you stumble on to the same estimate after 'data-mining', your (perhaps identical) theory should be neither supported nor rejected. Objectivity in any science should require that the identity of the investigator not affect the significance of an empirical observation. If the regression in question happened to be an exact replica of the process which generated the actual economic data, it would be ludicrous to accept it when it was arrived at by a first lucky guess and not when it was obtained by trial and error.

Apriorism is deeply ingrained in recent economic thinking. Yet many economists would not wish to adopt so strong a view as the new classicals seem to require. It is difficult to insist emphatically on the priority of economic theory when it is appreciated that there is no unanimity among economic theorists. Most, though not all, economists agree that economic theory should be grounded in a Walrasian general equilibrium approach founded on optimization by individual economic agents. There is less agreement on the use of the rational expectations hypothesis, although it is a common assumption. The consensus over what constitutes an adequate basis for economic theory does not rest on overwhelming empirical support.[7] Indeed, the strong apriorist denies that such support is, forthcoming. Rather, the consensus rests upon the tacit agreement of theorists trained and working in a particular framework. Peirce (1957: chapter 5, esp. p. 196) calls such tacit agreement within a community the 'method of public opinion' for fixing belief. He goes on to observe, however, that beliefs fixed by public opinion within a community frequently come unstuck when there is contact with another community with conflicting beliefs. In economics, communities may be defined by their purposes or the circumstances in which they work. It is well known how the changes in the purposes of economic policy in the face of stagflation in the early 1970s and the apparent failure of macroeconometric model to predict the course of economic events broke the consensus in macroeconomics and spawned alternative approaches of which new classicism was the most prominent theoretically. The new classical strategy in the face of the Lucas critique and the events of the 1970s is not to give up the search for invariance, but to broaden the scope of the search, and to found it on strong a priori principles. Two principles are fundamental: that agent's knowledge is systemic – i.e., it is based on, at least, implicit understanding of the true structure of the economy (rational expectations); and that agents are continuous and successful (to a serially uncorrelated random deviation) optimizers (Lucas and Sargent 1979: 304–9). Imposing such principles results in models for which the consequences of policy changes can be derived. Such models can then be empirically tested.

The difficulty with this strategy is that it provides neither guidance on how to proceed nor leeway to adjust assumptions, if the data are widely at variance with the model's predictions. The new classical organizing principles are uncontradictable.[8] Such principles are either not binding, because observationally identical results are just as well generated from other principles, or too constraining, because they rule out adjustments to the model which might reflect possible, but perhaps 'irrational' behavior. The problem with strong apriorism as a research strategy is principally that it does not naturally lead to a progressive development of models or knowledge. And, in fact, new classicism has not been able to offer a compelling alternative device for making economic predictions to the battered but stalwart community of macromodellers. For however great the influence of new classicism, and especially of the rational expectations hypothesis, on economic theory, it has not delivered any decisive empirical results.

As Cooley and LeRoy recognize, the thorough-going optimization implicit in new classicism suggests that everything depends on everything else, not just in theory, but in practice. Such complete interdependence suggests that theory does not provide sufficient restrictions to identify structural models. This shows up in the problem of observational equivalence: theoretical models which are antithetical nevertheless imply identical observational consequences (e.g., Sargent 1976). In the face of such a difficulty, Cooley and LeRoy retreat to theory, giving up hope of securing identification unless; there is a theoretical breakthrough. Others have given up theory and asked, 'what can be learned from the data alone?'[9]

The weak apriorist goes to neither of these extremes. Instead he recognizes that belief and inference stand in a relationship of mutuality: inferences are founded partly on unexamined beliefs; but these inferences, in turn, may suggest the modification of those beliefs. Thus, theory presents us with some a priori (in the sense of not *currently* questioned) restrictions on empirical investigation; while the empirical results help generate beliefs (or new theories) which are prior to further investigations. Haavelmo expresses the essentials of weak apriorism clearly:

> How can we actually distinguish between the 'original' system and a derived system . . . ? That is not a problem of mathematical independence or the like; more generally, it is not a problem of pure logic, but a problem of actually *knowing something* about the real phenomena, and of making realistic assumptions about them. In trying to establish relations with a high degree of autonomy we take into consideration various *changes* in the economic structure which might upset our relations, we try to dig down to such relationships as actually might be expected to have a degree of invariance with respect to certain changes in structure that are 'reasonable'. (Haavelmo 144: 29)

The objection to the new classical strategy of strong apriorism is not that it involves non-empirical principles (beliefs) – all empirical research does that. Rather, it is that it is committed so strongly to these beliefs that it does not permit them to adjust in the interplay of theorizing with the testing of theories. The strength of weak apriorism is precisely that it recognizes the need for such interplay.

Econometrics as Observation

The strong apriorist sets a formidable task for empirical economics. The title of Koopmans's attack on atheoretical economics suggests that econometrics, following a strict reading of its etymology, is concerned with the direct measurement of structures suggested by economic theory to be replicas of economic reality. But such a task is, as the weak apriorist suggests, not practicable unless we already have a good idea of what constraints reality places on the structures that we wish to measure. Measurement requires prior theory; equally, theory requires prior measurement.

That this circle seems vicious is the result of the apriorist failing to observe a critical distinction. Haavelmo distinguishes between *autonomous* relations, which are invariant to a wide range of interventions, and (adapting a term from Ragnar Frisch) *confluent* relations, which are the result of (complex) interactions of autonomous relations. Confluent relations may appear stable until subjected to interventions. Haavelmo's 'autonomous relations' are essentially the same as what the econometricians Hendry and Richard (1982) call the *data-generating process*: that is, the true, but unknown and unobservable description of how the data came to take on particular values.

Drawing this distinction forces us to recognize that only on the merest chance would we estimate a relation structurally identical to the data-generating process. Only on such a chance would we directly measure the underlying reality. And, what is more, since nothing in our estimate certifies its truthlikeness, we would never be completely sure that we had estimated the data-generating process. In general, we must assume that we estimate confluent relations.

Regressions and other econometric results are, first and foremost, calculations, summaries of observable data. Economists customarily speak of these calculations as 'good', 'bad', 'valid' and 'invalid'. Given the distinction between the data-generating process and confluent relations, it would be more to the point to think of these econometric calculations themselves as observations of the confluent relations. As such they may be illuminating or useful or neither, but not valid or invalid.

An analogy may make the point clearer. Astronomers use telescopes to observe the planets. The observations made with telescopes are not valid or invalid, but in focus or out of focus and, therefore, useful or not useful. The standard by which the usefulness of an astronomical observation is to be judged varies with what one seeks to observe. Filters that allow ultraviolet light to be singled out may wreck the normal visual spectrum. Whether or not this is good depends upon whether one wants to see ultraviolet and not green or blue. Econometric calculations are the economist's telescope, and the restrictions implicit in the specification of a regression, for example, act like the ultraviolet filter. The observations made with econometric calculations are observations of confluent relations, the consequences of the (probably) unknown data-generating process. As such they are the grist for the mill of theory. They are what theory must explain. Theory may in turn suggest new restrictions on econometric calculations as likely to be more illuminating than the initial ones. The ideal theory, nevertheless, explains not only these new results, but *all* observations – that is, it encompasses them. The strong apriorist view that econometrics should measure the coefficients of the data-generating process is clearly untenable. It amounts to the same thing as suggesting that astronomers directly observe Newton's laws with their telescopes, rather than the complex consequences for the planets of those laws.

The view adopted here that econometrics is best thought of as the observation of confluent relations shades into weak apriorism. It differs mainly in that it distinguishes sharply, as the weak apriorist does not, between the unobservable, but ultimately constraining, data-generating process and the observed confluent relations. Haavelmo, for example, treats autonomy in econometric relations as a matter of degree. Similarly, Zellner identifies law-likeness with invariance to a broad range of circumstances and boundary conditions.[10] This may be a good standard for *law-likeness*, but on the present view an actual *law* is distinguished in kind from such empirical relations by being an element of the data-generating process itself.

The analogy between econometrics and observational sciences such as astronomy suggests that criteria are needed to determine when an econometric calculation will be useful. That is, rules for focusing the telescope are needed. Such rules themselves are derived from theory – for the most part theory that both is not currently under scrutiny and is supplemental to the main investigation. Thus, a crude rule for focusing a telescope might be that the edges of the object in view should be sharply defined. This follows from the theory of optics that light travels in straight lines and from the presupposition that the object in view is in fact solid. Were the more central astronomical theory to suggest that the object in view was a gaseous cloud with poorly defined edges, optical theory might in turn suggest maximizing the received light as a focusing criterion. Theory, therefore, may modify observational strategy. On the other hand, it may have been the impossibility of observing sharp edges which suggested the modifications to the theory that generated such a change in strategy. The process is mutual. It would be absurd to give the central theory as dominant a role as strong apriorism demands. This would be equivalent to requiring Galileo to have had a theory of lunar geology before accepting that he had observed what we now know to be mountains on the moon. Theory governs our interpretation of what we observe; but its absence does not prevent us from observing something that needs interpretation and explanation.

Statistical theory provides the econometric equivalent of the focusing rule for the telescope. The basis for widely accepted criteria is the verisimilitude of an econometric specification or model with the data-generating process.[11] Not knowing the actual process, we can nevertheless say that a model cannot resemble it unless its errors are random – that is unless the part that we cannot explain is, at least provisionally, unexplainable, the model cannot be called truthlike. Typical criteria for randomness are: estimated errors should be white noise (i.e., not correlated with their own past – equivalently, they should have no autocorrelation); errors should be innovations (i.e., not correlated with other variables omitted from the model); and errors should be homoscedastic (i.e., of constant variance). If errors do not possess these properties, then it should be possible to formulate a different model that is better in the sense of having a lower variance and encompassing the first model. (In this case *encompassing* means providing a basis for calculating what the coefficients and variance of the other model would be without in fact estimating it.) In addition, on weak assumptions, statistical theory leads us to expect errors to be approximately normally distributed.

Just as in astronomy, theory may also guide econometric observations. At the crudest level, theory suggests potential variables. It also requires that models not imply values out of the range of possible observations. A consumption function, for instance, must not generate predictions of negative consumption. On a higher plane, we may impose restrictions from economic theory on econometric estimates and test these restrictions against more general models. If they are accepted, then theory aids in understanding the significance of the observation; if not, the observation may suggest what element of the theory is unsatisfactory.

Another requirement is stability of coefficient estimates. Like consistency with theory, this criterion is on a different plane from the need for random errors. The very concept of randomness – unexplainability – justifies it as a necessary condition of verisimilitude. There is no necessary connection between stability and the true data-generating process. Economic reality no doubt changes – perhaps, even so frequently that stability is not to be found. Nevertheless, econometric observations would be practically useless if they were completely unstable. We must, therefore, count on finding some stability and on supplementing econometric observations with other information, say institutional facts, if we are to distinguish between real changes in structure and our own inability to focus our observations. The criterion of consistency with theory is subject to similar strictures. It is useful only in that it aids interpretation of observations. We must not join strong apriorism in affording it overarching status. Observations must give grounds for reconsidering theoretical commitments.

4. Realism versus Nominalism

In adopting the view that econometrics is an observational tool, we take what is best from weak apriorism, while avoiding the pitfalls of strong apriorism. Our research strategy is progressive, because our observations are always known to be provisional, subject to improvement on grounds of better statistical technique or better theoretical interpretability, and because our commitment to a particular theory is not so strong as to preclude modification in the face of observations.

In adopting econometrics as observation we also implicitly commit ourselves to metaphysical realism, abjuring the nominalism implicit in strong apriorism. By nominalism I mean the philosophical doctrine that only individuals are real and that general relations (e.g., causality) do not exist independently of the observer.[12] The desire to avoid metaphysics is strong among most economists, having been brought up on the philosophy of logical positivism. But metaphysics cannot be avoided; it can only be done well or badly (cf. Peirce 1957: 53, 292, 293). A thorough discussion of nominalism and realism would take us too far down a philosophical byway. The nominalism of apriorist econometricians nonetheless presents a practical difficulty for their analysis.

Simon (1953: 24–6) noticed that any set of data could be represented by many incompatible causal structures. Such non-uniqueness is not a property of Simon's representation alone. Indeed, Sims (1977) shows in a set-theoretic analysis that a wide class of formal representations suffers from a similar lack of uniqueness. The most common view among econometricians is that the problem of non-uniqueness can be avoided only by a priori commitments to certain restrictions on allowable representations.

This problem arises from nominalism, which fails to distinguish between the relation as it exists in the world and the representation of it or the operational means of inferring it. So long as causality is seen as a relation which exists only because it is imposed by the observer, few restrictions will be placed upon acceptable formalisms; and even these few will arise in a priori commitments, not from observed reality.

Realism is the contrary doctrine to nominalism: namely, that general relations exist independently of the observer 'in the objects'.[13] Realism is the foundation of the conditional analysis of causality. For without it, counterfactuals do not make sense. The nominalist Mach writes: 'The universe is not twice given, with an earth at rest and an earth in motion; but only once with its relative motions, alone determinate. It is accordingly, not permitted us to say how things would be if the earth did not rotate' (Mach 1941: 284; cf. Simon 1952: 56). The realist objects that as we perfectly understand counterfactual claims, the nominalist is in no position to forbid them. The world *is*, in a sense, twice given. It is because we assume that certain causal relations exist in the objects that we are justified in making predictions. If we can predict on the basis of our understanding of what causal relations are, then we can equally well say what would have happened had antecedents been different. The world is twice given in the sense that our representation of the causal relations in it is not an arbitrary categorization, but a better or worse replica of the actual causal relations in the world. Hence, a counterfactual claim is sustained if it can be deduced from our representation (model or theory) on the assumption that antecedents are different from what they were in fact. The counterfactual claim, then, stands or falls with the satisfactoriness of our model or theory at truly representing the world. Even though such claims are simply formal deductions from our models or theories, they are nonetheless claims about the real world, not about our models.

Strong apriorism, either as an answer to the invariance problem or as a guide to good econometrics, adopts the nominalistic position that only particular facts are real and that general relations are not there to be found by the shrewd observer, but are imposed from without. It is the extraordinary view that theory is paramount and binds reality, rather than reality placing constraints upon what an acceptable theory would look like. On the strong a priori view, deductions from a theory may be disconfirmed, but such disconfirmation does not touch the theoretical core of the theory; it merely suggests that the optimization problem was not fully specified. The strong apriorist believes that economic observations are secure only if they are guided by a priori theory; but what is it that is supposed to make theory secure if it is not economic observation?

Econometrics as observation, in contrast, is grounded in realism. The econometrician does not impose restrictions but tries to learn in what way data constrain theories. The distinction between the model as a more or less good replica and the world is maintained. And the autonomy of the causal structure of the world is not questioned. The problem is not to decide how the world must be to agree with theoretical principles, but to discover how it is in fact.

The difficult problem of how to tell whether or not our theory is a good replica remains, but it is separate from the problem of whether causal relations are our own creations or are in the objects.

University of California, Davis

Notes

This paper is a revised version of Working Paper No. 33 of the Research Program in Applied Macroeconomics and Macro Policy, University of California, Davis. I am grateful for the comments of Thomas Mayer, Peter Oppenheimer, Edward Leamer, Thomas Cooley, Stephen LeRoy, and Steven Sheffrin.

1. Hoover (1988b). Mackie's analysis is also explored in Addison *et al.* (1980a,b) and Hammond (1986).
2. I say that the conditional proposition 'sustains' our inference, following Mackie, and that it is 'correct' rather than 'true' to avoid becoming directly embroiled in sharp debate between philosophers over how or whether truth values can be assigned to conditional propositions (see Hoover, 1988b: 6, esp. fn 7).
3. Generalizing beyond the linear even to very general mappings between variables is relatively easy; see Mesarovic (1969), Katzner (1983): chapter 6 and Hoover (1990): 232–4.
4. A point rediscovered by Buiter (1980), who observes that policy non-invariance requires only that agents take *some* account of policy rules, not that they have rational expectations.
5. For example, Johnston (1972): sections 12.2–12.4; the source for most textbook treatments of identifiability is Koopmans (1950).
6. A variation on this theme is the pre-test estimator, which penalizes the statistical significance levels according to the amount of search engaged in; see Judge *et al.* (1980): chapter 3 and Leamer (1978): chapter 5.
7. This proposition is amply documented for general equilibrium theory in Weintraub (1983).
8. This is admitted for the market-clearing assumption, see Lucas and Sargent (1979): 310–12.
9. E.g., Sims (1980). Despite the unsatisfactory nature of their own response, Cooley and LeRoy's criticism of the equivocations and invalid deductions drawn from vector autoregressions by Sims and others remains correct and important.

10. In keeping with its empiricist spirit, Engle *et al.* (1983): 285 identifies strong exogeneity (weak exogeneity plus invariance) with Zellner-causality (predictability according to law).
11. Fuller explanations and defence of these criteria are found in the writings of Hendry and his colleagues; see, e.g., Hendry and Richard (1982), Hendry (1983), Ericsson and Hendry (1984).
12. Goodman and Quine (1947) is a classic statement of the nominalist point of view; while Peirce (1957): chapters 1, 2, 4 and (1934): chapter 10, argues strongly against it, in favor of realism.
13. This phrase is Hume's (1888 [1739]: 88; see Mackie 1980: chapter 1). The philosophical doctrine of realism is expounded by Peirce (see n. 1, above) and presented as a basis for the philosophy of science by Newton-Smith (1981).

References

Addison, John T., Burton, John and Torrance, Thomas S. (1980a) 'On the causation of inflation', *Manchester School* 48 (2): 140–56.

——— (1980b) 'On the causation of inflation: some further clarifications', *Manchester School* 49 (4): 355–6.

Buiter, William (1980) 'The macroeconomics of Dr. Pangloss: a critical survey of the new classical macroeconomics', *Economic Journal* 90 (357), 34–50.

Cooley, Thomas F. and LeRoy, Stephen F. (1985) 'Atheoretical macroeconometrics: a critique', *Journal of Monetary Economics* 16 (3): 283–308.

Engle, Robert F., Hendry, David F. and Richard, Jean-Francois (1983) 'Exogeneity', *Econometrica* 51 (2): 277–304.

Ericsson, Neil R. and Hendry, David F. (1984) 'Conditional econometric modelling: an application to new house prices in the United Kingdom', unpublished typescript, Nuffield College, 31 August.

Goodman, Nelson and Quine, W. V. O. (1947) 'Steps towards a constructive nominalism', *Journal of Symbolic Logic* 12 (4): 105–22.

Granger, C. W. J. (1980) 'Testing for causality: a personal viewpoint', *Journal of Economic Dynamics and Control* 2 (4): 329–52.

Haavelmo, Trgyve (1944) 'The probability approach in econometrics', *Econometrica* 12 (supplement, July).

Hammond, J. Daniel (1986) 'Monetarist and antimonetarist causality', *Research in the History of Thought and Methodology* 4: 109–26.

Hendry, David F. (1983) 'Econometric modelling: the "consumption function" in retrospect', *Scottish Journal of Political Economy* 30 (3): 193–220.

Hendry, David F. and Richard, Jean-Francois (1982) 'On the formulation of empirical models in dynamic econometrics', *Annals of Applied Econometrics*, 1982–3 (supplement to the *Journal of Econometrics on Model Specification*), Halbert White (ed.) 20: 3–33.

Hoover, Kevin D. (1988) *The New Classical Macroeconomics: A Sceptical Inquiry*, Oxford; Basil Blackwell.

——— (1990) 'The logic of causal inference', *Economics and Philosophy* 6 (2): 207–34.

Hume, David. (1888 [1739]) *A Treatise of Human Nature*, Oxford: Clarendon Press.

Johnston, J. (1972) *Econometric Methods*, second edition, New York: McGraw Hill.
Journal of Econometrics (1988) 'Causality' Annals 1982–3 (supplement), Dennis J. Aigner and Arnold Zellner (eds), vol. 39.
Judge, George G., Griffiths, William E., Hill, R. Carter and Lee, Tsoung-Chao (1980) *The Theory and Practice of Econometrics*, New York: John Wiley.
Katzner, Donald W. (1983) *Analysis Without Measurement*, Cambridge: Cambridge University Press.
Koopmans, Tjalling C. (1947) 'Measurement without theory', *Review of Economics and Statistics* 29 (2): 161–72.
——— (1950) 'When is an equation complete for statistical purposes', in Koopmans (ed.) *Statistical Inference in Dynamic Economic Models*, New York: John Wiley; London: Chapman & Hall, pp. 393–409.
Leamer, Edward E. (1978) *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, New York: John Wiley.
——— (1985) 'Vector autoregressions for causal inference?', in Karl Brunner and Allan H. Meltzer (eds) *Understanding Monetary Regimes*, Carnegie-Rochester Conference Series on Public Policy, vol. 22.
Lucas, Robert E., Jr (1976) 'Econometric policy evaluations: a critique', in Karl Brunner and Allan H. Meltzer (eds) *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, vol. 1. Amsterdam: North Holland, pp. 19–46.
——— (1981) *Studies in Business-Cycle Theory*, Oxford: Basil Blackwell.
Lucas, Robert E., Jr and Sargent, Thomas J. (1979) 'After Keynesian macroeconomies', reprinted in Lucas and Sargent (1981), pp. 295–320.
Lucas, Robert E., Jr and Sargent, Thomas J. (eds) (1981) *Rational Expectations and Econometric Practice*, London: George Allen & Unwin.
Mach, Ernst (1941) *The Science of Mechanics*, La Salle, Ill.: Open Court.
Mackie, J. L. (1980) *The Cement of the University: A Study in Causation*, Oxford: Clarendon Press.
Mesarovic, Mihajlo D. (1969) 'Mathematical theory of general systems and some economic problems', in H. W. Kuhn and G. P. Szegö (eds) *Mathematical Systems and Economics I*, Berlin: Springer-Verlag, pp. 93–116.
Newton-Smith, W. H. (1981) *The Rationality of Science*, London: Routledge & Kegan Paul.
Peirce, Charles S. (1934) *Collected Papers*, vol. 6, Cambridge, Mass.: Belknap Press.
——— (1957) *Collected Papers*, vol. 7, Cambridge, Mass.: Belknap Press.
Sargent, Thomas J. (1976) 'The observational equivalence of natural and unnatural rate theories of macroeconomics', reprinted in Lucas and Sargent (1981), pp. 553–62.
Simon, Herbert A. (1952) 'On the definition of the causal relation', reprinted as chapter 3 in H. A. Simon (1957) *Models of Man*, New York: John Wiley.
——— (1953) 'Causal ordering and identifiability', reprinted as chapter 1 of Simon (1957), *Models of Man*, New York: John Wiley.
Sims, Christopher A. (1977) 'Exogeneity and causal ordering in macroeconomic models', *New Methods of Business Cycle Research: Proceedings of a Conference*, Minneapolis: Federal Reserve Bank of Minneapolis, pp. 23–44.
——— (1980) 'Macroeconomics and reality', *Econometrica* 48 (1): 1–48.
——— (1982) 'Policy analysis with econometric models', *Brookings Papers on Economic Activity* (1): 107–52.

Weintraub, E. Roy. (1983) 'On the existence of a competitive equilibrium: 1930–1954', *Journal of Economic Literature* 21 (1): 1–39.

Zellner, Arnold A. (1979) 'Causality and econometrics', Karl Bruner and Allan H. Meltzer (eds) *Three Aspects of Policy Making: Knowledge, Data and Institutions*, Carnegie-Rochester Conference Series oh Public Policy, vol. 10. Amsterdam: North-Holland, pp. 9–54.

SEVENTEEN

Does Macroeconomics Need Microfoundations?

Kevin D. Hoover

As I observed in the first lecture, I chose Pissarides's model as a paradigm of the modern macroeconomic model for a variety of reasons: the clarity of its goals and exposition; the manner in which it attempted to relate its theoretical construction to empirical facts (at least in principle); and, by no means the least important reason, because it was the model that Nancy Cartwright held up as an example of a nomological machine in economics. A number of fellow economists, however, question whether Pissarides's model really is a *macroeconomic* model. Because it appears to model the decision problem of the individual worker and the individual firm, some economists regard it as a microeconomic model. But this is all the better for my purposes because there is a persistent refrain in recent macroeconomics that the only acceptable macroeconomic models are those that have adequate *microfoundations*.

The idea of microfoundations did not originate with the new classical macroeconomics, but the manner in which the new classical macroeconomics has dominated the agenda of macroeconomics over the past quarter century has firmly cemented it in the minds of virtually all economists. Lucas puts it clearly when he longs for an economics that does not need the prefixes "micro" or "macro" – sound economics is held to be microeconomics, and any macroeconomics that is not just a shorthand for the manner in which microeconomics is applied to certain problems is held to be bad economics.[1]

Lucas advocates the euthanasia of macroeconomics and has spent most of his career supplying pills to hasten the demise of the once proud models of the macroeconomic era. It has taken time, but we have reached the point at which there are graduate students for whom John Hicks's IS/LM model

Pages 57–89 of *The Methodology of Empirical Macroeconomics*, by Kevin D. Hoover.

is just a dim memory from an undergraduate textbook and whose first lecture in their graduate macroeconomics courses began with a Hamiltonian describing the dynamic optimization problem of what appears to be an individual agent. Gradually, undergraduate textbooks are following suit, and even the econometric forecasting models of the United States Federal Reserve System have undergone surgery to remove the IS/LM model that once was the beating heart of their more than two hundred equation system. That the profession has sworn allegiance to the ideal of microfoundations is beyond doubt. The question before us is whether they are right to do so.

Some History

The earliest empirical economics is macroeconomics. The word "economics" derives from a Greek word meaning the management of the household. The earliest name for our subject, "political economy," consciously drew the analogy between the management of the household and the management of the state. But the politics of the seventeenth and eighteenth centuries was somewhat different from the politics of the nineteenth, twentieth, and twenty-first centuries. The transition to individualism was incomplete, and it was not uncommon for the political theorists of the day to think more of the social hierarchy as king, aristocracy, merchants, farmers, peasants, and so forth with little regard to the role of the individual. The early statistical researches of William Petty, Gregory King, and Charles Davenant were aimed not at understanding the economic behavior of particular people but at determining the capacities of England and Ireland to support the military ambitions of the English king. The models of François Quesnay and the Physiocrats, which bear many structural and methodological resemblances to modern macroeconomic models, went a step further. Again, appealing to the division of French society into broad classes (nobility, farmers, artisans), they gave normative advice to the French king on how to direct the economy in a manner that would enlarge his military capabilities.

The macroeconomic models of the seventeenth and eighteenth centuries were not supplanted all at once in a wave of individualism. The seeds had to be planted. The beginning of wisdom was the notion promoted by Adam Smith and the great Scottish political economists that the source of social welfare was the individual welfare of the ordinary man. We are so used to the idea that economics is about harnessing individual self-interest for social harmony and to attributing this idea to Smith, that we forget how limited were his claims for individualism. We remember the "Invisible Hand," but this image appears only once in the *Wealth of Nations*, in the context of foreign trade (and in two other instances in Smith's noneconomic works). Bernard Mandeville, early in the eighteenth century, in *The Fable of the Bees*, put the point that private vice (greed) could promote public virtue far more clearly than did Smith. But Smith took a dim view of Mandeville. Smith, David Ricardo, and the other classical economists were mainly concerned with market phenomena, and the individual played a relatively weak analytical and empirical role in their arguments.

With marginalism in the middle of the nineteenth century, the analytical ground shifts more clearly to the individual, but market phenomena remain the focus of William Stanley Jevons and the English political economists. It is in the work of the French economists Augustin Cournot and Leon Walras that the individual is truly made the analytical center of economics and the problem of how individuals coordinate socially, usually ascribed to Smith, takes center stage.

The political philosophy of the late nineteenth century is marked by debates over the relative explanatory role of individualism versus superindividual categories. Marxists led the way. For them, classes determine men, rather than men determining classes. (Yet, one should note that Karl Marx's economics owed its analytical framework to Smith and Ricardo and so was tainted, at least as far as they went with it, by individualism.) Austrian economics presented a clear contrast in which Carl Menger and, in the twentieth century, Ludwig von Mises, Friedrich von Hayek, and others espoused *methodological individualism*: the doctrine that the only well-grounded explanations of social phenomena were ones that appealed to the actions and behaviors of individuals.

English and American economics maintained an incomplete individualism. Although Alfred Marshall managed to kill the "political" that had long modified "economy" in the name of our discipline, his object was more to refocus attention on the analytics of the subject rather than on the applications. (The term "political economy" has been reborn in the past twenty years, though it conveys a very different sense now than it did in Smith's time.) Marshall discussed the particular firm and the particular worker or consumer. But, like his English and Scottish forefathers, he did so mainly to illuminate markets. The analyzed individual is meant to typify individuals in general. It is to Marshall, with his discussion of the "representative firm," that we owe the idea of the *representative agent*.[2] Still, Marshall's markets are not economy-wide, but are focused on particular products. Economics by 1930 appears mainly to be microeconomics. Yet, the proto-macroeconomics of the earlier time did not completely vanish. It is clearly evident in theoretical discussions of money, especially of the quantity theory, which never succeeded in finding adequate grounding in individual analysis. And it is evident in empirical discussions of business cycles, which were regarded as economy-wide phenomena.

So things stood in the mid-1930s, when John Maynard Keynes was writing the *General Theory*. Keynes did not invent macroeconomics, nor did he use the term. (As far as I can discover, Ragnar Frisch was the first to use the term, in 1931, though it became current only after the Second World War.)[3] Keynes, nevertheless, clarified the distinction between what we now call macroeconomics and microeconomics and made it possible for us to ask the question, how are the two related? As is evident in his discussion of the consumption function (the marginal propensity to consume follows from a "fundamental psychological law"), investment (entrepreneurs optimize with respect to opportunity costs), and the demand for money (speculators anticipate capital gains or losses), Keynes follows Marshall in looking to the individual decision problem for illumination. These appeals to individual behavior remain in the service of aggregate explanations. Despite the fact – largely ignored in potted histories – that he stresses the heterogeneity of individual responses as a central feature of aggregate behavior, Keynes never explores the relationship between the individual and the aggregate in any really systematic way.

Microeconomics so dominated economic thinking in 1936 that the cry for microfoundations for the newly resurgent macroeconomics was almost immediate. Jacob Viner and Wassily Leontief wrote microeconomic criticisms of the *General Theory*.[4] Lawrence Klein, in his *Keynesian Revolution*, thought it necessary to discuss the microeconomic underpinnings of the principal Keynesian aggregate functions.[5] The history of the first twenty-five years of postwar macroeconomics is largely the hanging of micro-economic flesh on the skeleton of interpretation of Keynes's *General Theory* formalized in Hicks's aggregate general-equilibrium, IS/LM model. James Dusenberry, Milton Friedman, and Franco Modigliani tried to explain the microeconomics of consumption; William Baumol and James Tobin, the demand for money; Dale Jorgenson, investment; Don Patinkin, labor; and so forth.[6]

Beginning with Robert Clower's dual-decision hypothesis and Robert Barro and Herschel Grossman's fixed-price models, the urge for microfoundations began to infect the general-equilibrium framework.[7] It is no longer enough that each function have an individualistic foundation; since individuals are assumed to be making choices to generate each function separately, those choices really ought to be coordinated and consistent. This is a hard problem with heterogeneous agents. The modern representative agent, which is essentially a homogeneity assumption, made his appearance first in these models. At more or less the same time, Lucas and Leonard Rapping began to model unemployment as an optimization problem. Lucas made consistent optimization in general equilibrium the centerpiece of his monetary model published in the *Journal of Economic Theory* in 1972.[8] Strictly speaking, this model is not a representative-agent model. Yet, it is highly idealized and assumes that all individuals are fundamentally identical. From there, it is only a short step to the representative-agent models that have dominated new classical macroeconomics since the early 1970s.

Reductionism

So much for a brief history of the movement for micro-foundations in economics. What are the intellectual roots of this urge to ground macroeconomics in the individual? It has analogies in other sciences. The nature of scientific explanation is a hotly debated subject among philosophers and scientists. One plausible view is that a theory is explanatory when it achieves parsimony: if a complex phenomenon can be reduced to some smaller number of governing principles, then we regard the complex phenomenon as having been explained.

In the eighteenth century the ideal gas laws were formulated. The Boyle-Charles law states that

$$pV = nRT,$$

where p is pressure, V is volume, n is the number of moles of the gas, R is the universal gas constant, and T is temperature. As the name suggests this law is an idealization of the results of empirical observations and holds with a high degree of accuracy at moderate temperatures and low pressures.

The gas law appears to be an approximate truth about physical reality, but nevertheless physicists were not happy with its sui generis quality. The solution is found in the kinetic theory of gases, which provides an account of the gas laws as a deduction from Newtonian mechanics. The kinetic theory is also based on an idealization: the gas is assumed to be composed of molecules regarded as perfectly elastic point masses. With the added assumption that the velocities of the molecules are distributed according to a particular random distribution – that they are equally likely to move in every direction – it is possible to derive the gas laws. Temperature corresponds to the mean energy of the molecules and pressure to the mean momentum transferred by contact with the walls of the containing vessel. The kinetic theory of gases thus constitutes a *reduction* of the macrophysical gas laws to the microphysical Newtonian mechanics.

Notice two features of this reduction. The first is that it is not micro all the way down. In addition to Newton's laws, the kinetic theory relies on a statistical assumption – that is, an implicitly macro assumption. Also, notice that the categories that apply to Newton's laws and to the gas laws are very different. A single molecule has momentum and energy, but it does not have pressure or temperature. To make the derivation work, it is necessary to identify aggregate properties of the collection of molecules (their *average* energy and momentum) as corresponding to the macro properties (temperature and pressure) that have quite different sensible characteristics. The phenomena of temperature and pressure can be thought of as *emergent properties* of the aggregation of molecules.

Reductionist strategies are pursued throughout science. Recently, in biology, a lot of effort has been directed to reducing macrobiological phenomena to the micro principles of genetics and organic chemistry. But even here, the effort is controversial, with one wag saying: "the only way to reduce biology to chemistry is through death."[9] The philosophical mind/body problem has, in the age of neuroscience, also generated a debate over reductionism. The issue is whether mental states can be completely explained by knowledge of brain states. Even if they could, the issue of the phenomenological difference between the two levels is larger here than it is with respect to the gas laws. Seeing a beautiful woman does not seem to be the same kind of thing as any pattern of neuron firings. Vision and, to a greater degree, aesthetic appreciation appear to be emergent properties, even if there is a reduction.

The situation is even more complex than that. You and I can see the same thing even though our brain states are not the same. Similarly, you can see the same thing at different times even though your brain state is different at each time. There is no one-to-one mapping between the macro phenomena of mind and the micro phenomena of brain states. This observation has led to the notion of *supervenience.* Mental states are said to supervene on brain states in the sense that any time one could exactly reproduce a certain brain state and collateral conditions, the same mental state would occur, even though that mental state may occur for other configurations of brain states as well, and even though the appropriate phenomenological descriptions of the mental state are completely different from those of the brain states. Supervenience guarantees the autonomy of the macro level in the sense that it ensures that one can rationally use an independent language and categories to describe the macro level and that one should not expect to find unique deductions from the micro to the macro. Yet, it also underscores the connection between the micro and the macro: no macro state exists unless an appropriate micro state exists. Supervenience has been offered both as a way of eliminating the need for reduction and as a justification for a weaker form of reduction. Which way one looks at it partly depends on what one views as threatened.

Economics and Methodological Individualism

So what about reductionism in economics? Whether economic explanations must be reductive depends in part on how one defines economics. An older tradition defines it with respect to certain areas of human life. The classic definitions can be summarized in a word: *plutology*, the science of wealth. John Stuart Mill writes:

> Writers on Political Economy profess to teach, or to investigate, the nature of Wealth, and the laws of its production and distribution: including, directly or remotely, the operation of all the causes which the condition of mankind, or of any society of human beings, in respect to this universal object of human desire, is made prosperous or the reverse.[10]

Similarly, Alfred Marshall writes:

> Political Economy or Economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing.
>
> Thus it is on the one side a study of wealth; and on the other, and more important side, a part of the study of man.[11]

Modern economists almost all follow the much different definition of Lionel Robbins:

> Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses.[12]

Economics is, in Robbins's view, the science of choice. Economics is, in modern terminology, microeconomics.

Once microeconomics is seen as defining the very nature of economics, any macroeconomic phenomenon will be seen to need a reductive explanation. Of course, it is one thing to want such an explanation and quite another to have it. It is obviously impractical to dispense with measurements of temperature and pressure and to keep track of the velocities of each and every molecule even in a relatively small volume of gas. Similarly, it is absurd to think that practical economics can trace the decisions and constraints facing each individual agent in the economy. I call this the *Cournot problem*, because the first clear statement of it is found in Cournot's

Researches into the Mathematical Principles of the Theory of Wealth (1838). No one really denies the Cournot problem; the only question is what to do about it.

Notice that the motivations for seeking a reduction are different in economics than they are in biological sciences. Biologists are suspicious, for instance, of mental explanations because they involve *intentional* states: beliefs, purposes, desires, will, goals, and so forth. Human mental life is *teleological*; that is, it is directed to ends. The reduction of the mental to the neurological is appealing to scientists precisely because neurons, chemicals, molecules, genes, and such do not have ends or intentional states. Reduction banishes teleology. In economics, it is just the reverse. Macroeconomic relations, say as represented in Okun's law, which relates changes in the unemployment rate to the growth rate of gross domestic product (GDP), are not obviously intentional anymore than the gas laws are. But if macroeconomic relations are regarded as the products of human action, this could be seen as a defect. The goal of reducing macroeconomics to microeconomics is to recapture human intentions. Reduction reclaims teleology.

The difference is clear in what is probably the most influential paper in macroeconomics in the postwar period: Lucas's "Econometric Policy Evaluation: A Critique."[13] Lucas criticized the empirical macroeconomics of the day – especially the large-scale macroeconometric forecasting models – on the basis that their equations captured transitory correlations in the data that would not remain stable in the face of changes in policy regimes. His idea is that people make choices subject to constraints that include their best expectations of government policy. If the government uses the macroeconomic models to guide its policy choices, it will surely find that the models fail as soon as it changes its policy, because agents will adapt to the constraints of the new policy. Projecting macroeconomic relationships estimated in the past into the future implicitly assumes that the policy of the past continues. But if the government uses those projections to guide changes in its policy, then it assumes that people expect the old policy, even while a new policy is in place. People are not stupid, so the past projections are bound to fail. The most common response to the Lucas critique (for example, in the program of Lars Peter Hansen and Thomas Sargent and, more recently, in real-business-cycle models) was to argue that economic projections would be secure only if they were grounded in a deep analysis of the decision, problems faced by individuals, including their detailed understanding of the structure of policy.[14] A model was said to be secure from the Lucas critique only if it was grounded in relationships built up from the "deep parameters" corresponding to tastes and technology. Only well-specified optimization problems were supposed to provide a secure basis for economic prediction. In other words, macroeconomics must be reduced to microeconomics. The conviction that macroeconomics must possess microfoundations has changed the face of the discipline in the last quarter century.

That the argument for microfoundations should have been so successful rhetorically is, I think, puzzling. For it ignores the obvious difficulties in empirical implementation posed by the Cournot problem. As I said before, no one believes that economists can practicably trace the decision problems of millions of individuals and aggregate them to discover macroeconomic behavior. The intellectual triumph of microfoundations is grounded not in methodological individualism (that is, in a strategy of basing all empirical explanations on the behavior of individuals) but in *ontological individualism* (the conviction that the only real entities in the economy are individuals). Who could disagree with that?

Well, I would. Unfortunately, the full argument for this position would take us further down a metaphysical byway than any group of economists is likely to want to go. Still, I would at least like to poke a few holes in the presumption that ontological individualism is necessarily correct. The fear of the ontological individualist is that if he says that macroeconomic entities like GDP or the general price level are real, he must also say that they are independent of the individual people who constitute the economy. The second claim is, of course, obviously wrong, but ontological individualism does not follow from denying it.

The relationship between microeconomics and macroeconomics could be one of supervenience. Any identical reconfiguration of the agents in the economy and their situations results in the same configuration of the macroeconomic entities in the economy, but the mapping is not one to one. What is more, the supervenience of the macroeconomy on the microeconomy is not just a weak form of reductionism. This is because of intentionality at the microlevel. Individuals have to make plans and decisions on the basis of expectations about the future. In so doing, they face precisely the same problem that is faced by the economist from his detached perspective: the economy is too complex for a detailed microeconomic account to inform the construction of expectations. Individuals, just like economists, face the Cournot problem. When I try to figure out how much money to put aside to pay for my daughters' college education, I must make guesses about future inflation and interest rates, as well as about my own income. I cannot do that by constructing a realistic computable-general-equilibrium model of the economy. Instead, I use simple macroeconomic models (indeed, crude time-series models, such as one that says that future interest rates will be the average of past interest rates). But this means that I cannot completely reduce macroeconomics to microeconomics. Microeconomics of the real world necessarily uses macroeconomic models and concepts as an input. The macroeconomy supervenes on the microeconomy but is not reducible to it.

Aggregation and the Illusion of a Microeconomic Ontology

While I am convinced that the impulse that made the microfoundational argument succeed is ontological and not methodological, it would be absurd not to acknowledge the methodological sea change in macroeconomics after the Lucas critique. Macroeconomic models look like microeconomic models (hence the reaction that my use of Pissarides's model provoked among my colleagues). The same techniques, the same mathematics, the same language is used. But this is truly puzzling. The physicist who has successfully reduced the ideal gas laws to the kinetic theory of gases does not then abandon the language of pressure, temperature, and volume when working with gases or try to use momentum, mass, and velocity as the principal phenomenological categories for discussing the macroscopic behavior of gases.

But economists have taken a different tack. They have typically started with the microeconomics of the individual and then asked to what degree the lessons learned at that level can still apply to aggregates of individuals. There is, in consequence, a vast literature on the theory of aggregation. The general conclusion of this literature is that aggregation in which the macro looks like the micro can occur only under circumstances so stringent that they could never be fulfilled in the real world except by the merest chance. I want to argue something even stronger than that; namely, that even what appears to be perfect aggregation under ideal circumstances fails. But, first, let us consider the lessons of aggregation theory as they stand.

Economics is about heterogeneous things. In microeconomics we choose how to allocate our consumption among different goods or how to allocate factors of production used to make those goods. In both cases, we consider physical things of disparate natures and somehow have to make them equivalent. The role of utility functions or profit functions is to give us a common denominator, a basis for choosing among goods that otherwise are little alike. Similarly, when we calculate nominal GDP, we cannot add up the disparate goods until we have given them a common denominator – typically, money. Real GDP is even one step further removed, as we correct the monetary unit of measurement for changes in its own value by constructing a notion of a *general* price level.

Now, the first question asked in aggregation theory is, when is aggregation perfect? – that is, when can two disparate goods be added together and treated analytically as if they were but one good? The criteria are typically economic, not physical, though the first example may seem physical. Suppose that we have a certain quantity of coal and a certain quantity of oil. Coal and oil differ on many dimensions; but, if the only difference of material importance to us is the amount of heat they produce (which dimensions matter is the economic criterion), then we can measure each in British Thermal Units (BTUs), rather than in tons or barrels, and add them up in those units. This is the case in which, up to a factor of proportionality, the goods are perfect substitutes. Similarly, in any case in which goods are perfect substitutes on the relevant dimensions, we can aggregate them.

Oddly, the polar opposite case works as well. Consider the manufacture of water through burning hydrogen and oxygen. It takes exactly two moles of hydrogen and one mole of oxygen to make one mole of water. We cannot vary the formula. Hydrogen and oxygen are not substitutable; they are perfect complements. But we can aggregate perfectly by counting bundles of hydrogen and oxygen into bundles: 2H + 1O = 1 water bundle.

Generally, however, except in these extreme cases, perfect aggregation is not possible. The reason is economic. If goods are neither perfect complements (in which case no change in the mix of the goods is possible) nor perfect substitutes (in which case no change in the mix of goods matters), then the mix of goods can be changed and still yield the same output or utility. How that mix changes depends on relative prices. As the price of a good rises, we purchase less of that good and more of its substitute. This is the basis for the common claim, going back to Hicks, that we can treat bundles of goods as composite commodities, so long as their relative prices do not change: the so-called *composite commodity theorem.*[15]

The composite commodity theorem is true as far as it goes, but notice how special are the assumptions on which it is based. We generally regard prices not as exogenous variables given outside the economic system, but as one of the important products of economic coordination. The proofs of the existence of a general equilibrium, going back to Kenneth Arrow and Gerard Debreu, demonstrate that there is a set of prices that coordinates economic activity. The prices are not themselves parameters, but change as the true parameters (tastes and technology, if we go back to Lucas's formulation) change. The composite commodity theorem, therefore, holds only when the relevant underlying parameters do not change. How relevant can that be for interesting economic analysis?

Let us illustrate the problem with an extremely simple example. Consider an economy with two consumers and two goods. These goods can be either two goods in a single period or one physical good that can be consumed in two different periods. It does not matter which interpretation we take for the example to work, although the second one is directly relevant to a number of intertemporal macroeconomic models. Let each individual (i) choose the goods (c_1 and c_2) by maximizing a Cobb-Douglas utility function:

$$u^i = \log c_1^i + \alpha^i \log c_2^i \tag{17.1}$$

subject to a budget constraint

$$y^i - c_1^i - pc_2^i = 0, \tag{17.2}$$

where y is exogenously given income, and p is the price of good 2 in terms of the numeraire, good 1. The demand for good 1 is

$$c_2^i = \frac{y^i}{1+\alpha^i}. \tag{17.3}$$

Letting the superscripted, lower-case letters designate variables that apply to individual agents and upper-case or unsuperscripted letters, variables that apply to aggregates, the idea of the representative-agent model is simple. If equation (17.3) gives the demand for the individual for good 1, then the aggregate demand for good 1 is

$$C_1 = \frac{Y}{1+\alpha}. \tag{17.4}$$

But, in our simple economy of only two agents, it is easy to check exactly what the aggregate form, of the demand for good 1 should be. It is merely the sum of the two individual demands, so that

$$\begin{aligned} C_1 = c_1^1 + c_2^1 &= \frac{y^1}{1+\alpha^1} + \frac{y^2}{1+\alpha^2} = \frac{(1+\alpha^1)y^1 + (1+\alpha^2)y^2}{(1+\alpha^1)(1+\alpha^2)} \\ &= \frac{Y + \alpha^1 y^1 + \alpha^2 y^2}{(1+\alpha^1)(1+\alpha^2)}, \end{aligned} \tag{17.5}$$

since $Y = y^1 + y^2$. In general, equation (17.5) does not have the same form as equation (17.4). In fact, the only circumstances in which (17.4) and (17.5) are identical in form is when $a^1 = a^2 = a$ – that is, when all agents have identical tastes.

As a rule, the conditions are even more stringent than that. I purposely chose a very tractable utility function. The Cobb-Douglas utility function is homothetic; that is, its indifference curves are each parallel blowups of the indifference curves closer to the origin. Equivalently, the income-expansion paths (that is, the locus of tangencies between indifference curves and budget constraints as the budget constraint is moved outward to reflect increasing income and constant relative prices) are all straight lines through the origin. And this is what the theorists tells us: some technical details and caveats to one side, perfect aggregation from individual agents to a representative agent requires that all agents have identical utility functions and that these be homothetic. Why? Because in these cases, income distribution is not relevant. Because of homotheticity, the ratios of goods consumed by any one individual remain the same whether that individual is rich or poor. And because utility functions are identical, the ratios of goods consumed are the same for any individual. In such circumstances, for a fixed aggregate income, redistributing that income among the individual consumers will not affect demands for individual goods and, therefore, will not affect relative prices. In that case, the conditions of Hicks's composite commodity theorem apply, and we can add up individual quantities to form economy-wide aggregates without loss of information.

Although the example that we have looked at is extremely simple, it carries a very general message. The conditions of exact aggregation are strong and almost certainly never fulfilled in any practical instance. Why should one accept the representative-agent model and the facile analogy from the micro to the macro? Indeed, recently, a number of economists – Rolf Mantel, Hugo Sonnenschein, and Debreu – have shown that theoretically there is no such analogy.[16] No matter how well behaved the microeconomic functions may be, the aggregate functions, given distributional variations, are essentially unrestricted and need not take a form that is derivable in any simple way from the form of the underlying micro functions. This means, for example, that if every underlying production function is Cobb-Douglas, there is no theoretical reason to conclude that the aggregate production will also be Cobb-Douglas. Conversely, if the aggregate production function for an economy is Cobb-Douglas (which to a first approximation it appears to be for the U.S. economy), there is no reason to believe that this tells us anything at all about the shape of the underlying production functions.

There is a strong belief, expressed not only in the ordinary practice of macroeconomics but in the methodological writings of philosophers of economics, that aggregation does not alter the fundamental categories of economics. Whereas in physics molecules have one sort of description and gases, even though they are aggregations of molecules, quite another, in economics real GDP is much like any other real good. Uskali Mäki makes the point I wish to oppose by saying that economics does not add to the "ontic furniture" of the world given to common sense.[17] This is, I think, an illusion that arises because of the view that perfect aggregation represents a possible limiting case of actual aggregation. The possibility of perfect aggregation suggests the analogy of real GDP to an individual good. If, for example, relative prices are constant (that is, P_j/P_k is constant for all j and k), then $\Sigma_{j=1}^{n} P_{j,t} Q_{j,t}$ (where the t in the subscript indicates the base time, period t) can be normalized by choosing the units for the $Q_{j,t}$ so that each $P_{j,t} = 1$. Then, nominal GDP at time n can be written

$$\sum_{j=1}^{n} P_{j,t+n} Q_{j,t+n} = P_{t+n} \sum_{j=1}^{n} Q_{j,t+n}. \tag{17.6}$$

Under the assumed conditions P is unique. Some conclude, therefore, that in this limited case, one can treat the summation on the right-hand side of equation (17.6) as a natural aggregate quantity analogous to an individual quantity. The conditions for constant relative prices are almost certainly never fulfilled; but, even if they were, the summation is not analogous to an individual quantity. The general price level P in (17.6) still has the dimension period-n dollars/period-t (i.e., base period) dollars. To sum heterogeneous goods, they must still be converted to a common denominator, and in this case, the summation still has the dimensions of period-t dollars. This would be more perspicuous if (17.6) were written as

$$\sum_{j=1}^{n} P_{j,t+n} Q_{j,t+n} = P_{t+n} \sum_{j=1}^{n} 1_{j,t+n} Q_{j,t+n}, \tag{17.7}$$

where the subscripted numeral 1 is a place holder for the dimensional conversion.

One might regard perfect aggregation as the idealization of typical aggregation in which quantities are affected by changing relative prices. The upshot of the argument here is that the aggregate remains analogous to the macro gas of the ideal gas laws and is not obviously some natural extension of a single underlying molecule. The ideal gas laws fit well only within a limited range of temperatures and pressures. Outside that range, they, vary in a manner than can be accounted for using the kinetic theory of gases by adding more realistic assumptions about the volume of individual molecules and the forces acting between them. The equivalent in macroeconomics is found in the efforts of Alan Kirman and Kathryn Dominguez and Ray Fair, among others, to account for distributional effects in macroeconomic relationships.[18]

The Strange Career of the Representative-agent Model

Given what we know about representative-agent models, there is not the slightest reason for us to think that the conditions under which they should work are fulfilled. The claim that representative-agent models provide micro-foundations succeeds only when we steadfastly avoid the fact that representative-agent models are just as aggregative as old-fashioned Keynesian macroeconometric models. They do not solve the problem of aggregation; rather they assume that it can be ignored. While they appear to use the mathematics of microeconomics, the subjects to which they apply that microeconomics are aggregates that do not belong to any agent. There is no agent who maximizes a utility function that represents the whole economy subject to a budget constraint that takes GDP as its limiting quantity. This is the simulacrum of microeconomics, not the genuine article.

This seems transparently obvious. So why have intelligent economists come to believe so fervently both in the necessity of microfoundations and in the efficacy of the representative-agent model in providing them? Let me offer a speculation. One of the earliest examples of modern dynamic economics is found in Frank Ramsey's optimal savings problem.[19] In this problem, Ramsey considered the problem of saving for an economy and imagined it to be a social planner's problem in which the utility function represented social preferences, without conjecturing how these might be related to the preferences of the members of society. Ramsey may well have thought (in the manner of Keynes) that the wise men of Cambridge could be trusted to know what was best for society independently of any direct knowledge of the lower classes. Push-pin may have been as good as poetry for Jeremy Bentham; but Bentham was an Oxford man. In Cambridge the poets ruled and aspired to rule the world. On Cambridge assumptions, there is no problem with what Ramsey did.

By the early 1950s, the general-equilibrium model had been more thoroughly developed and analyzed. The two theorems of welfare economics were established:

1. Every perfectly competitive general equilibrium is Pareto efficient; and
2. Every Pareto-efficient allocation can be supported as a perfectly competitive equilibrium for some set of lump-sum transfers.

These two theorems appear to promise an isomorphism between social planner problems that choose Pareto-efficient allocations and perfectly competitive equilibria. In fact, this isomorphism provides a powerful technical tool for the solution of dynamic optimization problems, because it is often easier to define a social planner's problem and a Pareto-efficient outcome, and then to ask how to decentralize it, than it is to solve for the competitive equilibrium directly (a trick common in the literature on real-business-cycle models).

Notice that there is a sleight of hand here. Only rarely do macroeconomists care about the redistributions needed to decentralize the social planner's problem. It is fine to ignore redistributions when they do not matter – that is, when all agents are identical and have homothetic utility functions. Once again, the macroeconomists have slipped in unwarranted microeconomic assumptions, as well as, implicitly, assumptions about the shape of the social planner's function. But, if we take the notion of decentralization seriously, we know that everyone cannot be alike. Furthermore, not only does aggregation theory tell us that we do not know how the social planner's function might relate to the underlying utility functions, the older Arrow Impossibility Theorem tells us that, for reasonable assumptions, no social planner's function exists that respectfully and democratically aggregates individual preferences.[20] Thus, the idea of the representative agent appears to arise naturally in dynamic macroeconomic models as a kind of benign extension of Ramsey's social planner in the face of the two welfare theorems. But this idea is plausible only when the macroeconomist fails to take microeconomics seriously.

Could we, nevertheless, not regard the representative-agent model as an idealization? It may be a good way to think about macroeconomic problems when the losses due to aggregation are relatively small. Let us accept that, but notice that whether or not the representative-agent model is a good thing depends now entirely on its contingent empirical success. It may work; it may solve the Lucas critique; it may not. We just have to see. There is no longer a point of principle involved. The advocate of the representative-agent model has no right to attack other macroeconomists for failing to provide microfoundations, for he fails to provide genuine microfoundations himself.

My guess is that the representative-agent model may help in pointing to some sorts of qualitatively useful relationships. But it is unlikely to provide useful quantitative restrictions on the behavior of macroeconomic aggregates. The reason can be seen by thinking about the way in which Marshall used the idea of the representative, firm. For Marshall, the representative firm was not the average, or even median, firm, but a firm that typified firms at a point in their life cycle at which the extreme behaviors associated with very small or very young firms, on the one hand, or very large or very old firms, on the other hand, could be set aside. If we can analogize back to the physicist's ideal gas, Marshall wanted to describe the usual behavior of a gas molecule under certain ideal conditions. The use of representative-agent models in modern macroeconomics attempts something quite different. It attempts to describe the behavior of the gas (its pressure and volume), not by considering seriously how the molecules behave in aggregate, but by analyzing the gas as if it were one big molecule subject to the laws that in fact govern real molecules. This is a category mistake: pressure and volume are descriptions of the properties of aggregates – properties that individual molecules either in reality or idealized to colossal size do not possess as isolated units.

On the analogy with gases, we should conclude that what happens to the microeconomy is relevant to the macroeconomy but that macroeconomics has its own descriptive categories and may have its own modes of analysis. It is almost certain that, just as in the case of gases, no genuine microfoundations can ever be provided for macroeconomics that do not make concessions to the macrolevel in the form of statistical assumptions about the distributions of important microeconomic characteristics. And, given those concessions, it is almost certain that macroeconomics cannot be euthanized or eliminated. It shall remain necessary for the serious economist to switch back and forth between microeconomics and a relatively autonomous macroeconomics depending upon the problem in hand.

Suggested Readings

As observed in this lecture, the history of microfoundations is a long one. The modern obsession with microfoundations as the sine qua non of macroeconomics can be dated to Robert E. Lucas, Jr.'s "Econometric Policy Evaluation: A Critique" (originally published in Karl Brunner and Allan H. Meltzer [eds.], *The Phillips Curve and Labor Markets*, vol. 1 of Carnegie-Rochester Conference Series on Public Policy, Amsterdam: North-Holland, 1976, and reprinted in Lucas's own *Studies in Business Cycle Theory*, Oxford: Blackwell, 1981). An excellent methodological study of the necessity of microfoundations is found in Maarten Janssen's *Microfoundations: A Critical Inquiry* (London: Routledge, 1993).

More particularly, the modern ploy of providing microfoundations through the representative-agent model is brilliantly attacked in Alan Kirman's "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6(2) (1992), 117–36, and, with a rich historical perspective, in James Hartley's *The Representative Agent in Macroeconomics* (London: Routledge, 1997).

Notes

1. Lucas (1987), pp. 107–108.
2. Marshall's notion is, as we will see, substantially different from that common in modern macroeconomics; see Hartley (1996, 1997).
3. Frisch used the term in his lectures; Erik Lindahl may have been the first to use it in print in 1939; see Fitoussi and Velupillai (1993).
4. Viner (1936) and Leontief (1936).
5. Klein (1947).
6. Dusenberry (1949), Friedman (1957), Modigliani and Brumberg (1954), Baumol (1952), Tobin (1956, 1958), Jorgenson (1963), and Patinkin (1965).
7. Clower (1965) and Barro and Grossman (1971).
8. Lucas (1972).
9. Vercelli (1991), p. 243.
10. Mill (1848/1911), p. 1.
11. Marshall (1920), p. 1.
12. Robbins (1935), p. 16.
13. Lucas (1976).
14. Hansen and Sargent (1980).
15. Hicks (1946), p. 46.
16. Kirman (1992) and Hartley (1997).
17. Mäki (1996).
18. Kirman (1992), Dominguez and Fair (1991).
19. Ramsey (1928).
20. Arrow (1951).

References

Arrow, Kenneth J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Barro, Robert, and Herschel I. Grossman. (1971). "A General Disequilibrium Model of Income and Employment." *American Economic Review* 61(1), 82–93.

Baumol, William J. (1952). "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Quarterly Journal of Economics* 66(4), 545–56.

Clower, Robert W. (1965). "The Keynesian Counter-revolution: A Theoretical Appraisal." Reprinted in Donald A. Walker (ed.), *Money and Markets: Essays by Robert W. Clower*. Cambridge: Cambridge University Press, 1984, pp. 34–58.

Dominguez, Kathryn M., and Ray C. Fair. (1991). "Effects of the Changing U.S. Age Distribution on Macroeconomic Equations." *American Economic Review* 81(5), 1276–94.

Dusenberry, James. (1949). *Income, Saving and the Theory of Consumer Behavior*. Cambridge: Cambridge University Press.

Fitoussi, J. P. and K. Velupillai. (1993). "Macroeconomic Perspectives," in H. Barkai, S. Fischer, and N. Liviatan (eds.). *Monetary Theory and Thought*. London: Macmillan.

Friedman, Milton. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Hansen, Lars Peter, and Thomas J. Sargent. (1980). "Formulating and Estimating Dynamic Linear Rational Expectations Models." *Journal of Economic Dynamics and Control* 2(1), 7–46.

Hartley, James E. (1996). "Retrospectives: The Origins of the Representative Agent." *Journal of Economic Perspectives* 10(2), 169–77.

Hartley, James E. (1997). *The Representative Agent in Macroeconomics.* London: Routledge.

Hicks, John R. (1946). *Value and Capital,* 2nd ed. Oxford: Clarendon Press.

Jorgenson, Dale. (1963). "Capital Theory and Investment Behavior." *American Economic Review* 53(2), 247–59.

Kirman, A. P. (1992). "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspective* 6(2), 117–36.

Klein, Lawrence R. (1947). *The Keynesian Revolution.* New York: Macmillan.

Leontief, Wassily. (1936). "The Fundamental Assumption of Mr. Keynes's Monetary Theory of Unemployment." *Quarterly Journal of Economics* 51(1), 192–97.

Lucas, Robert E., Jr. (1972). "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4(2), 103–24.

Lucas, Robert E., Jr. (1976). "Econometric Policy Evaluation: A Critique." In Karl Brunner and Allan H. Meltzer (eds.), *The Phillips Curve and Labor Markets.* Carnegie-Rochester Conference. Series on Public Policy, vol. 1, Spring. Amsterdam: North-Holland, pp. 19–46.

Lucas, Robert E., Jr. (1987). Models of Business Cycles. Oxford: Blackwell.

Mäki, Uskali. (1996). "Scientific Realism and Some Peculiarities of Economics." In R. S. Cohen, R. Hilpinen, and Qiu Renzong (eds.), *Realism and Anti-Realism in the Philosophy of Science.* Dordrecht: Kluwer, pp. 427–47.

Marshall, A. (1920). *Principles of Economics: An Introductory Volume,* 8th ed. London: Macmillan.

Mill, J. S. (1988/1911). *Principles of Political Economy with Some Applications to Social Philosophy.* London: Longman's and Green.

Modigliani, Franco, and R. Brumberg. (1954). "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In K. Kurihara (ed.), *Post-Keynesian Economics.* New Brunswick, NJ: Rutgers University Press.

Patinkin, Don. (1965). *Money, Interest, and Prices.* 2nd ed. New York: Harper and Row.

Pissarides, Christopher A. (1992). "Loss of Skill during Unemployment and the Persistence of Employment Shocks." *Quartely Journal of Economics* 107(4), 1371–91.

Ramsey, Frank P. (1928). "A Mathematical Theory of Saving." *Economic Journal* 38(152), 543–59.

Robbins, L. (1935). *An Essay on the Nature and Significance of Economic Science.* London: Macmillan.

Tobin, James. (1956). "The Interest Elasticity of the Transactions Demand for Cash." *Review of Economics and Statistics* 38(3), 241–47.

Tobin, James. (1958). "Liquidity Preference as Behaviour Towards Risk." *Review of Economic Studies* 25(2), 65–86.

Vercelli, Alessandro. (1991). *Methodological Foundations of Macroeconomics: Keynes and Lucas.* Cambridge: Cambridge University Press.

Viner, Jacob. (1936). "Mr. Keynes on the Causes of Unemployment." *Quarterly Journal of Economics* 51(1), 147–67.

EIGHTEEN

Economics in the Laboratory

Vernon Smith

Vernon Smith (1927–) received his Ph.D. in economics at Harvard. Although his more than two hundred books and articles address issues in many areas of economics, he is best known for his work on experimental economics, for which he received the Nobel Prize in 2002. After many years at the University of Arizona, Smith is now a professor of economics at George Mason University.

Why do economists conduct experiments? To answer that question, it is first necessary briefly to specify the ingredients of an experiment. Every laboratory experiment is defined by an *environment*, specifying the initial endowments, preferences and costs that motivate exchange. This environment is controlled using monetary rewards to induce the desired specific value/cost configuration (Smith, 1991, 6).[1] An experiment also uses an *institution* defining the language (messages) of market communication (bids, offers, acceptances), the rules that govern the exchange of information, and the rules under which messages become binding contracts. This institution is defined by the experimental instructions which describe the messages and procedures of the market, which are most often computer controlled. Finally, there is the observed *behavior* of the participants in the experiments as a function of the environment and institution that constitute the controlled variables.

Using this framework of environment, institution, and behavior, I can think of at least seven prominent reasons in the literature as to why economists conduct experiments. Undoubtedly, there are more (Davis and Holt, 1992, Chapter 1 and passim).

1. Test a theory, or discriminate between theories. This motivation comes from the economic and game theory literature. We test a theory by

Journal of Economic Perspectives, vol. 8 (Winter 1994): 113–31. Reprinted by permission of the American Economic Association.

comparing its message or its outcome implications with the experimental observations. The greater the frequency with which the observations hit these "predictions," in the context of a design in which hits are unlikely to occur by chance, the better the theory.[2] Examples can be found in the auction literature (Smith, 1991, 25–29), where risk averse models of bidding in Dutch and first price sealed bid auctions are favored by the data over risk neutral models, while dominant strategy auctions such as the English, whose outcomes are predicted to be independent of risk attitude, perform well in the laboratory. Of course, theories subjected to sufficiently rigorous tests are nearly always found to need improvement; this leads to the second reason for doing experiments.

2. Explore the causes of a theory's failure. When the observations of an experiment fail to conform to the implications of the theory, the first thing to be done is to reexamine the design, and to be sure that the predictive failure is the fault of the theory. Well-articulated theories formally model the environment and the trading rules, and the experimentalist seeks to reproduce these conditions of the theory. In the course of testing, when the experimental design continues to seem appropriate and the theory still fails, this tends to encourage an experimental examination designed to discover the cause. Establishing the anatomy of failure is essential to any research program concerned with modifying the theory. Examples are to be found in the bargaining literature (Roth, 1987; Hoffman and Spitzer, 1985; Hoffman, et al., 1992; Bolton, 1991) and in common value auctions (Kagel and Levin, 1986; Cox and Smith, 1992). Often theories that initially perform poorly show improvement if subjects are given more experience (Cox and Smith, 1992), or the payoffs are increased (Smith and Walker, 1993), but sometimes these measures fail to yield results that improve the theory's performance (Smith and Walker, 1993).

3. Establish empirical regularities as a basis for new theory. Well-formulated theories in most sciences tend to be preceded by much observation, which in turn stimulates curiosity as to what accounts for the documented regularities. Microeconomic theory tends to build upon simplifying assumptions, and to eschew attempts to model many of the complex trading and contracting institutions that we observe. But in the laboratory, especially with computerization, institutions with complex trading rules are as easy to study as are simple single unit auctions. This makes it possible to range beyond the confines of current theory to establish empirical regularities which can enable theorists to see in advance what are the difficult problems on which it is worth their while to work. The continuous double auction, used the world over, is a fine example. In this institution, buyers announce bid prices, while sellers announce offers, or asking prices. Any new bid (offer) must be at a price which is lower (higher) than the standing bid (offer); that is, the bid-asked spread must narrow. A binding contract occurs when a buyer accepts a seller's ask, or a seller accepts a buyer's bid. Contracts occur in sequence as new bids, asks and acceptances occur. Because of its robust equilibrating properties with small numbers of traders possessing only private information, this institution (Smith, 1991, 1, 2, 6) was studied extensively in the laboratory long before the attempts by R. Wilson, D. Friedman and others to model it (see Friedman and Rust, 1992, for references).

4. Compare environments. Comparing environments using the same institution permits an investigation of the robustness of that institution. The objective is to stress the theory with extreme environmental conditions under which an institution's established properties may begin to break down. Thus, in common value auctions (where the item has the same value to all bidders after the auction is completed), the Nash model performs better when there are 3–4 bidders than when there are 6–7 bidders (Kagel and Levin, 1986). Similarly, the Nash equilibrium prediction performs fairly well in the Fouraker and Siegel (1963) bargaining environment, but breaks down in the ultimatum game environment (Hoffman et al., 1992), as discussed below.

5. Compare institutions. Using identical environments, but varying the market rules of exchange, has been the means by which the comparative properties of institutions has been established. Examples include the comparison of English, Dutch, first and second price sealed bid auctions, the comparison of uniform and discriminative price multiple unit auctions, and the comparison of posted (retail) pricing with double auction trading (Smith, 1991, 25, 5, 17).

6. Evaluate policy proposals. Friedman's (1960) original proposal that the Treasury auction securities in one-price auctions led to their comparison with the discriminative rules (Smith, 1991, 5). Bids to buy in this auction are arranged from highest to lowest; if the offering was \$2 billion worth of bills, this amount of the highest bids are accepted at a price given by the highest rejected bid. In the past decade, private industry and government sponsors have funded studies of the incentives for off-floor trading in continuous double auction markets, alternative institutions for auctioning emissions permits, mechanisms for allocating space shuttle resources, and market mechanism for the allocation of airport slots (Plott, 1987).

7. The laboratory as a testing ground for institutional design. A growing use of the laboratory is as a testing ground for examining the performance properties of new forms of exchange. The early experiments studying the one-price sealed bid-offer auction for Treasury securities helped Henry Wallich to motivate the Treasury in the early 1970s to offer some long-term bond issues using this procedure (Smith, 1991, pp. 511–12). This led eventually to the use of the procedure in auctioning commercial paper and in setting the dividend rate on variable rate preferred corporate securities. In 1992, Treasury resumed its earlier experiments with the one-price auction because of publicized irregularities in dealer bidding.

A second example is the new Arizona Stock Exchange (AZX). In 1988, we started running our first experiments with the uniform price double auction. In this mechanism, buyers submit bids to buy, and sellers submit offers to sell in real time during the specified market "call" period. All bids, offers, and the tentative market clearing uniform price are displayed as they are entered, so participants can see the existing state of the market, and alter their own bids or offers accordingly. It turns out that this approach has efficiencies comparable to those of continuous double auction, but with no price discrimination. Subsequently, we learned that Steven Wunsch independently developed a similar system, and was seeking SEC authority to operate it as a proprietary stock exchange for institutions. Wunsch Auction Systems opened in New York in 1991. About this time officials of the Arizona Corporation Commission, who had heard of our experimental studies of "electronic exchange," approached us with the idea of starting an Arizona Stock Exchange. We demonstrated the uniform price double auction for them, pointed out its properties, and they were eager to get moving. Our first action was to get them together with Wunsch to explore the possibility of moving his exchange to Arizona. Eventually, Wunsch adopted the new name, AZX, and the new exchange has experienced rapid growth since its move in March 1992. Had it not been for the experiments we would not have come to understand the comparative properties of the uniform price double auction, and would not have been able to recommend it wholeheartedly as a reasonable direction for a new electronic exchange.

What Have Economists Learned from Experiments?

Hoffman's (1991) "Bibliography of Experimental Economics" contains 1500 entries. I can only attempt to report a small selection of some of the findings.

Institutions Matter

Experimentalists have long known that the continuous double auction rules of trade in securities markets constitutes a mechanism remarkably adept at maximizing the gains from exchange at prices tending to converge to competitive equilibria (Smith, 1991, 1). What we have learned since is that this is just one of many illustrations of the principle that institutions matter. This is because the rules determine the information states and individual incentives in the trading game: institutions matter because incentives and information matter. Consequently, posted offer retail pricing converges more slowly and erratically and is less efficient than continuous double auction (Plott and Smith, 1978). Unlike the latter, sellers receive no continuous bid price information from competing buyers. Also, sellers must quote one price per period for all units making price cuts more costly.

Does this mean that posted offers are inferior to continuous double auction? No. The experiments evaluate only the allocative properties of the two mechanism, and do not address their different transactions cost properties. With continuous double auction, every trade involves decentralized multilateral negotiation, while pricing is centralized in a posted offer system, and clerks need have no bargaining skills. The latter is cost effective for mass retail distribution, the former has been well-suited to the broker-dealer structure of securities markets.

As early as 1965 (Smith, 1991, 4), an extreme environment was used as a stress-test to explore the limits of the ability of the continuous double auction to generate competitive equilibria. This was the "swastika" environment in which the demand price is constant up to a maximum quantity, and the supply price (below demand price) is also constant up to a maximum quantity greater than the maximum demand quantity. If you draw these demand and supply curves you see what looks like a swastika emblem. Such markets still performed efficiently, but convergence to the competitive equilibrium was slow and erratic when the excess supply was very small. Van Boening and Wilcox (1992) have recently reported a much more successful stress-test of continuous double auction. They report experiments in which the sellers' only costs are fixed costs that can be avoided by selling zero units, and the demand price is constant up to a fixed capacity. This lumpy environment is structured so that there is no uniform price competitive equilibrium like that to which continuous double auction usually converges; yet efficient allocations exist. The important result is that continuous double auction cannot handle this environment, and research is under way for new or traditional mechanisms that can handle such cases. The issue is of practical importance. Airlines, for example, have large flight costs that can only be avoided by not flying.

One of the better-known predictive failures of expected utility theory is the "preference reversal" phenomenon. A subject reports that gamble A is preferred to B, but in responding with her selling price places a higher price on B (say $10) than on A ($7) (Lichtenstein and Slovic, 1971). But Chu and Chu (1990) report that such reversals are much reduced on the second iteration of a process in which the experimenter arbitrages the inconsistency, and reversals disappear on the third iteration, establishing that subjects are not satisfied with their own choices when they experience the implications of those choices. More subtle experiments have been reported by Cox and Grether (1992), in which each subject's selling price is elicted in an English Clock auction which is known to have good demand revelation properties. In this auction a clock is set at a low price; all buyers respond with their demands. The clock then ticks up to successively higher prices, and buyers respond by reducing their demand until there is but one unit demanded. After five repetitions, subjects' selling prices were in general consistent with their choices. Consequently, this provides another example of the tendency for rational behavior to emerge in the context of a repetitive market institution. But in this case, the market corrects the inconsistency of behavior found in choice elicitation experiments.

Unconscious Optimization in Market Interactions

In his early path-breaking critique of the feasibility of rational calculation in human choice, Simon (1955, p. 104) explicitly did not "rule out the possibility that the unconscious is a better decision-marker than the conscious." Unknown to both of us at the time was the fact that the first of hundreds of continuous double auction experiments reported in Smith (1991, 1, Chart 1) would spotlight the crucial importance of not ruling out the rationality of unconscious decision in rule-governed repeat interaction settings. Consider the typical conditions of a continuous double auction experiment. Subjects have private information on their own willingness-to-pay or willingness-to-accept schedules which bound the prices at which each can profitably trade. No subject has information on market supply and demand. After an experiment, upon interrogation they deny that they could have maximized their monetary earnings or that their trading results could be predicted by a theory. Yet despite these conditions, the subjects tend to converge quickly over time to the competitive equilibrium. Thus, "the most common responses to the market question were: unorganized, unstable, chaotic, and confused. Students were both surprised and amazed at the conclusion of the experiment when the entrusted student opened a sealed envelope containing the correctly predicted equilibrium price and quantity" (Gillette and DelMas, 1992, p. 5).

That economic agents can achieve efficient outcomes which are not part of their intention was the key principle articulated by Adam Smith, but few outside of the Austrian and Chicago traditions believed it, circa 1956. Certainly I was not primed to believe it, having been raised by a socialist mother, and further handicapped (in this regard) by a Harvard education, but my experimental subjects revealed to me the error in my thinking.

In many experimental markets, poorly informed, error-prone, and uncomprehending human agents interact through the trading rules to produce social algorithms which demonstrably approximate the wealth maximizing outcomes traditionally thought to require complete information and cognitively rational actors.[3]

Information: Less Can Be Better

Providing subjects with complete information, far from improving market competition, tends to make it worse. In 1976, I reported continuous double auction results, using the "swastika" environment described above, comparing the effect of private with complete information (Smith, 1991, 6). Under private information, convergence to the equilibrium outcome (in this case, the Nash-competitive outcome) was much more rapid and dependable then under complete and common information.[4] Similar results had been reported earlier by Fouraker and Siegel (1963) for Bertrand and Cournot oligopoly, and more recently by Noussair and Porter (1992) and Brown-Kruse (1992). When people have complete information they can identify more self-interested outcomes than Nash (and competitive) equilibria, and use punishing strategies in an attempt to achieve them, which delays reaching equilibrium.

Of course, it can be said that all of this simply supports the "folk theorem" that repetition aids cooperation. But the folk theorem operates in situations with small numbers and complete information – like the fact that a repeated prisoners' dilemma game tends to converge to cooperation. The argument here is much stronger: competitive tendencies prevail under the private information conditions that pervade markets in the economy.

The principle that private payoff information can yield "better" results has also been established in the Nash bargaining game (Roth, 1987). Nash assumed that the bargainers knew each other's utilities (preferences). Roth and his coworkers implemented this theory with ingenious simplicity: subjects bargained over the division of 100 lottery tickets, each representing a chance to win fixed large or small prizes for each of the two players, with the prizes generally being different for the two players. When the two players know only their own prizes (and each other's percentage of the lottery tickets), the outcome conforms to the Nash bargaining solution. When the bargainers also know each other's prizes the Nash prediction fails; in short, Nash theory is not falsified, it is just not robust with respect to the bargainers knowing both prizes.

The principle that less information can be advantageous also applies under asymmetric payoff information in which Schelling (1957) argued that the less informed bargainer may have an advantage over a completely informed adversary. In fact, Siegel and Fouraker (1960) observed this to be the case. The better informed bargainer, knowing that the other player knew only his own payoff, is more forgiving when his opponent makes large demands. This concessionary posture works to the disadvantage of the completely informed player. Camerer, Loewenstein and Weber (1989) call this the "curse of knowledge" and report new evidence in a market setting.

Common Information Is Not Sufficient to Yield Common Expectations or "Knowledge"

It has been argued that game theory requires common knowledge.[5] This arbitrarily limits the value of game theory in organizing experimental data, and directs our attention away from the fact that we must understand the process of achieving common knowledge if game theory is to make progress in predicting behavior. This is implicitly recognized by the growing current interest among game theorists in concepts of bounded rationality and of learning. Although I believe these are conceptually the right directions to take, if the exercises are guided by introspective model development, uninformed by observations and testing, they are unlikely to achieve their full predictive potential.

Experimentalists have attempted to implement the condition of "common knowledge" by publicly announcing instructions, payoffs, and other conditions in an experiment. Some examples of this process would be Roth (1987) in Nash bargaining games, Smith, Suchanek and Williams (Smith, 1991, 19) in finite horizon asset trading experiments, and McCabe (1989) in finite horizon fiat money experiments, but there are many others. However, it should be noted that administering common instructions in public literally achieves common information – *not* common knowledge in the sense of expectations. In other words, there is no assurance that a public announcement will yield common expectations among the players, since each person may still be uncertain about how others will use the information.

In laboratory stock markets, each player receives an initial endowment in cash and shares of stock. It is public information that the expected dividend in a given time period will be some fixed number for each of the T periods of the game. With zero interest rate, the value of a share of stock in the first time period should be T times the expected dividend. In each time period, the rational expectations hypothesis is that share prices will be equal to the remaining dividends to be paid, and will decline by an amount equal to the expected dividend in each time period.

In fact, first-time participants in experiments of this sort – whether they are undergraduates, graduates, business persons, or stock traders – produce bell-shaped price bubbles starting below fundamental value, rising well above and crashing to near fundamental value in the last few periods. Trading volume is high. When subjects return for a second session, the price bubbles are dampened, and volume is reduced. When they return for a third session, trading tends to follow the decline in fundamental value, with very thin volume. These experiments illustrate that participants come to have common expectations by experience, not by being given common information and then reasoning that others will expect prices to be near fundamental value.[6]

Unless players have common expectations of behavior in later periods, they cannot reason backwards to the present. This problem, for theories based on backward induction rationality, is illustrated by the wage search experiments of Cox and Oaxaca (1989). In their experiments, subjects search a distribution of wages, and must decide in each period whether to accept a certain wage offer; if accepted the subject must forego continued search and the possibility of receiving a better subsequent offer. In this situation, subjects have only to anticipate their own behavior in later periods in order to practice backward induction. Subjects in these experiments behave as if they are solving the backward induction problem properly. Hence, it would appear that when common expectations exist (because the subject "knows" his or her own expectations) then subjects will backward induct. Of course, this does not mean that subjects are conscious of having solved such a problem, and can tell you about it.

In this journal, Brandenberger (1992) has usefully emphasized that the assumption of common knowledge is sometimes unduly strong; examples are given in which if each of two players are rational and they have mutual knowledge (both know it, but not necessarily that both know that both know it), then a Nash equilibrium follows. These distinctions between various degrees of knowledge are certainly helpful, but if game theory is to have *predictive* value, it is necessary to go further and seek to discover *operationally* how to achieve the required conditions of knowledge. Theories based upon abstract conditions make no predictions. Subjects obtain knowledge of the strategy choices of others by experience. This is why I see no way for game theory to advance independently of experimental (or other) observations. We have to understand the processes whereby the required conditions of knowledge are satisfied – processes like pre-game play, repeated play, cheap talk, or the futures market example discussed in a moment – before the implications of those conditions can become testable hypotheses.

It has been observed that if the failure of rational expectations in finite horizon trading experiments was due to the lack of common expectations about later periods, then introducing futures markets should hasten convergence to rational expectations equilibria by speeding up the process of creating common expectations of later period behavior (Porter and Smith, 1989). Forsythe, Palfrey and Plott (1982) had reported that convergence in two-period horizon experiments was hastened by introducing a futures market on period two. If our interpretation was correct, then a futures market on period eight in 15-period asset trading experiments would aid in creating common expectation at mid-horizon (subjects already expect trading at fundamental value near the end), and price bubbles should be retarded in the presence of such a futures market. Porter and Smith (1989) report experiments supporting this hypothesis. The learning suggested by these studies is that the important role of futures markets may be to foster common expectations among traders concerning a future event. This permits the backward induction calculus to yield the appropriate rational expectations in the current period.

Dominated Strategies Are for Playing, Not Eliminating

It is commonly argued that dominated strategies should never rationally be played, and thus can be eliminated in game-theoretic analysis. But players in repeated games do sometimes play dominated strategies and there are sound reasons why.

Consider the two-person alternating-play game tree in Figure 18.1, which is played repeatedly for a long time with uncertain termination (McCabe, Rassenti and Smith, 1992). If player 1 moves down at x_1 then at x_2 player 2 can signal a desire to achieve the cooperative outcome (50 : 50) by moving left, or, by moving right, signal a desire to achieve the subgame noncooperative outcome (40 : 40), since player 2 knows that player 1 will see node x_6 as more attractive than node x_4. But if player 2 chooses left at x_2, player 1 can defect by moving down at x_3, forcing player 2 at node x_5 to choose between (60 : 30) and the direct punishment outcomes that result at node x_7. Game theory reasons that player 2 should play left at node x_5, accepting player l's defection, but punish on the next round of repeated play by choosing right at node

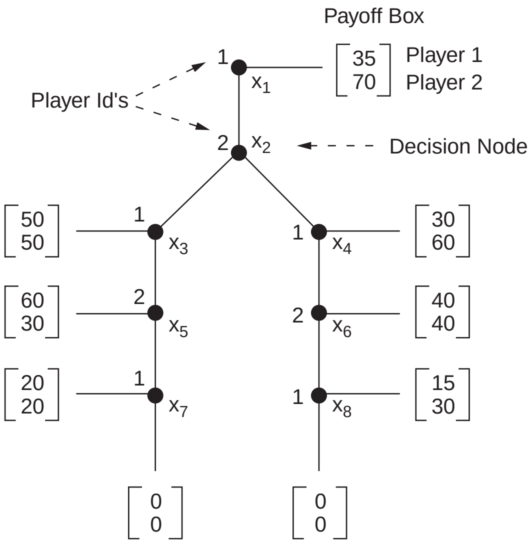

Figure 18.1. A Two-Person Alternating-Play Game

x_2 (choosing right at x_2 almost without exception ends at the equilibrium (40 : 40)). Subject player 2's tend not to do this, but instead to play down at x_5, and thereby to punish immediately. The reason is clear, the resulting message is unambiguous, with no possibility that player l's will misunderstand. The strategy works: even when 12 subjects are randomly repaired after each play, there is a strong tendency toward the cooperative outcome by round 15–20. (If the game is altered by interchanging the (50 : 50) and (60 : 30) payoff boxes, thereby removing player 2's ability to punish immediately, then cooperation fails to emerge). This is not the game-theoretic route to repeated-play cooperation because the bargainers are assumed to have common expectations (knowledge). But, as we have seen, common expectations is achieved by a process of play, not by deductive analysis. Part of this process may be to punish in ways that will be clearly understood.

Efficiency and Underrevelation Are Compatible

It is well-known that a market participant, whether a buyer or seller, can sometimes tilt the conditions of the transaction toward personal gain and away from market efficiency, by not revealing true willingness to trade. Consequently, economists often seem to argue as if market efficiency must rely on complete revelation of preferences.

As an empirical counterexample, consider the version of the uniform price double auction mechanism studied in McCabe, Rassenti and Smith

(1992). Remember that in this auction format, subjects submit openly displayed bids during a market call period. In this format, subjects greatly underreveal demand and supply, but they adjust their bids and offers so that the market clearing price and quantity approximates a competitive equilibrium. At this equilibrium, they produce many bids and offers tied at the same price. This behavior serves to protect each side of the market against manipulation by the other side. That is, if a buyer attempts to lower the market price by bidding lower, that buyer's bid is replaced by another tied bid without moving the price, and similarly if a seller attempts to raise the price.

In short, efficiency only requires enough revelation to allow the marginal units on both sides of the market to trade. This can occur although there is massive under-revelation of the inframarginal units. In uniform price experiments, one frequently observes that subjects capture 100 percent of the surplus while revealing only 10–15 percent of it in their bids.

The Endowment Effect

Thaler (1980) has argued that the observed tendency in survey studies for willingness-to-accept to exceed willingness-to-pay by nontrivial amounts is due to an "endowment" (or ownership) effect which arises because of loss aversion; an example is the man who paid $5 per bottle for a case of wine. A few years later he is offered $100 per bottle, and refuses, although he has never paid more than $35 for a bottle of wine. In this case giving up the wine yields a loss which is more highly weighted than the gain from purchasing an equivalent bottle. The existence of an endowment effect has been suggested by numerous hypothetical survey studies; recently, the experimental focus has been to verify its existence with real goods.

It has been argued by Kahneman, Knetsch and Thaler (1990) that the endowment effect does not apply to goods held for resale; only to goods which are consumed. Similarly, it does not apply to the exchange of rights (or tokens) on which value has been induced by cash payments in experiments. In either case, since what is being acquired is intended from the start to be resold, losses and opportunity costs are transparently equivalent. Kahneman, Knetsch and Thaler (1990) report both choice and exchange experiments confirming the results with tokens, but establishing the willingness-to-accept/willingness-to-pay discrepancy for consumer goods (like emblem mugs, pens, and so on). They also reject empirically the important qualification that the discrepancy is due to income effects (see their experiments 6 and 7). Franciosi, Kujal, Michelitsch and Smith (1993) have reported experiments that narrow the reported willingness-to-accept/

willingness-to-pay discrepancy by using a more uniform choice task, and by using the uniform price double auction (with its good revelation properties for marginal units) to establish price. While these results reduce the discrepancy, the endowment effect remains statistically (and economically) significant.

Samuelson and Zeckhauser (1988) suggest that the endowment effect may be a manifestation of a broader "status quo bias;" they provide results showing the existence of such a bias even when the problem is not framed in terms of gains and losses. Models of utility-maximizing when decision costs are taken into account postulate a trade-off between the sum of all the various costs of decision-making and the value of the decision outcome (Smith and Walker, 1993). Such models predict a bias in favor of one's current status, since any change is cognitively and information costly.

Fairness: Taste or Expectation?

According to survey studies reported by Kahneman, Knetsch and Thaler (1986), people indicate that it is unfair for firms to raise prices and increase profits in response to certain changes in the environment which are not justified by an increase in costs. Thus, respondents report that it is "unfair" for firms to raise the price of snow-shovels after a snowstorm, or to raise the price of plywood following a hurricane. In these circumstances, economic theory predicts shortages, an increase in prices toward the new market clearing levels, and eventually an increase in output. However, economic theory does not predict the verbal behavior of agents in this process so such expressions do not falsify the theory.

Do expressions of unfairness reflect interpersonal utilities that reduce effective demand for the product of offending parties, or do they vent the unpleasant need for expectations to be adjusted? If such results show no more than a lag as aggrieved parties adjust their expectations to the new reality, the standard models will predict the eventual result, as the indignation subsides. But protesting parties may react strategically in their self-interest by withholding demand and punishing price "gougers," or, fearing this, sellers may moderate or forgo their increase in prices. Alternatively, by way of contemporary contract theory, one side or the other may see the reference price and transactions as an implicit contract, not to be lightly tampered with. If an economic agent can extract resources by claiming unfair treatment, then it is consistent with standard theory for the agent to manufacture words to that effect. In such situations, it isn't clear that standard self-interested utility-maximizing models can account fully for the observed market behavior.

Kahneman, Knetsch and Thaler (1986) do not predict the final outcome in these cases; a departure from the reference transaction, initially seen as unfair, may eventually achieve the status of a new reference transaction. This argument is a form of the standard adaptive expectations hypothesis, and has been tested in an experimental market environment (Kachelmeier, Limberg, and Schadewald, 1991; Deng et al., 1992). In an initial baseline series of trading periods with a 50 percent profits tax on sellers, the after-tax profit of sellers is identical with the consumer's surplus of buyers, and the division of surplus is "fair." Then the reference baseline is altered by substituting a 20 percent sales tax for the 50 percent profit tax on sellers. The effect of the sales tax is to raise the market clearing price, and substantially increase seller after-tax profit relative to buyer profit in comparison with the reference situation. Across experiments, the subjects are divided into three different treatment groups: (1) marginal cost disclosure, in which buyers are informed of the price implications of the sales tax; (2) no disclosure, in which buyers are given no new information; (3) profit disclosure, in which buyers receive a graph showing for each price what the potential split of total surplus is between buyers and sellers.

Deng et al. (1992) choose a particular institutional context in which sellers independently post selling prices at the beginning of each period. Buyers, queued at random, choose to make their purchases one at a time. The Kahneman, Knetsch, Thaler argument implies that in the first period, prices will be highest under marginal cost disclosure, where buyers are informed of the price implications of the sales tax, because the disclosure serves to justify price increases and to reduce any resistance to them. Revealing profits, on the other hand, will lead to the lowest prices in the first period, because the change from the reference (baseline) transactions is greatest, and will lead to substantial resistance. The no-information group should, according to the hypothesis, fall between these extremes.

The results strongly and significantly support the Kahneman, Knetsch and Thaler hypothesis. In period one, the price in the marginal cost disclosure group was very near the new competitive equilibrium, with prices much the lowest in the profit disclosure group. But in successive trading periods, the mean prices in the profit disclosure and no disclosure groups increase, and by period 10 none of the three means are statistically different from each other or the competitive price. These results offer strong confirmation of standard theory, as the sellers in the profit disclosure treatment raise prices over time in response to the excess demand. Furthermore, as sellers raise prices they are not deterred by any significant incidence of demand withholding by buyers.

Fairness questions also arise in the ultimatum game where a sum of money, say $10, is to be allocated between two people. Player 1 moves first offering some amount, *X*, of the $10 to player 2. If player 2 accepts that amount, then player 1 receives the rest; if player 2 rejects that amount, both players receive zero. Game theory predicts that player 1 will offer the smallest possible amount to player 2; player 2 will accept it as better than nothing; and player 1 will take the lion's share. However, in the experimental context when players are anonymously paired, and play only once, the modal offer by player 2 is $5, with a lower median.

These observations have been interpreted as showing that the players have a taste for fairness (see Bolton, 1991, and his references). In particular player 2 is concerned about being treated fairly by player 1, and the latter must take this into account lest her offer be rejected. But this interpretation has been called into question by the results of the "dictator game" in which player 2 *must* accept the offer of player 1. Forsythe et al. (1988) find significantly lower offers in the dictator game than in the ultimatum game. Hoffman et al. (1992) corroborate these results and report dramatically lower offers (two-thirds offer zero) when the dictator game is run double blind: the experimenter does not know the decisions or payoffs of any subject. To put it another way, the dictator results are highly sensitive to the degree of anonymity from other persons. This suggests that the ultimatum game results are due primarily to strategic and expectational considerations, and not just to a taste for fair outcomes. The same considerations apply to the above market experiments.

Methodology and Experiment

The fact that the planet Mercury exhibited an orbit that violated Newton's theory did not lead Newtonians to conclude that the theory was falsified; rather, they concluded that there must exist a heretofore unknown planet between the sun and Mercury that perturbed its orbit from the predicted path. They even named it Vulcan, and there was no subsequent shortage of claimed sightings (Roseveare, 1982). *All* tests of a theory require various auxiliary hypotheses that are necessary in order to interpret the observations as a test of the theory. These auxiliary hypotheses go under various names: initial conditions, *ceteris paribus* clauses, background information, and so on. Consequently, all tests of a theory are actually joint tests – that is, a test of the theory conditional on the auxiliary hypotheses. This leads to the Duhem-Quine theses, according to which one can always rescue a theory from an anomalous observation by *ex post hoc* recourse to imaginative and persuasive auxiliary hypotheses. Conversely, every observational victory for a theory can be questioned by a suitable revision of the background knowledge in which the theory is embedded. This thesis denies the possibility of direct falsification of any specific testable implication of a theory (and, in its strong form, denies rational rules of selection).

My view is that some philosophers have exaggerated the significance of the Duhem-Quine problem, while experimentalists may be unaware of its power in influencing their day-to-day activities. Experimental economists are intuitively if not formally aware of the problem; this is why they do so many experiments probing the sources of a theory's failure, or success, as in the ultimatum game and other examples discussed above. If you have a confounding problem with auxiliary hypotheses, then you do new experiments to test them. If the auxiliary hypotheses are not testable, this is preeminently your critic's problem.

A recent exchange among experimentalists in the December 1992 *American Economic Review* is squarely reflective of the Duhem-Quine problem. Harrison (1992) has questioned all falsifying observations in experimental economics as due to a postulated low opportunity cost of deviating from theoretical optimality. This thesis sets the stage for the convenient nihilist belief that all recalcitrant observations must be due to inadequate payoff opportunity cost. (Of course, this argument raises the unanswered question of why there exist validating results with low opportunity cost). But, like most important instances of Duhem-Quine, the proposition can be and has been tested – in this case many times over the last 30-odd years (Smith and Walker, 1993, offer a review).[7] The results have made it plain that money does matter; that factors besides money also matter; that many anomalies do not disappear by escalating payoffs (and foregone profits); and that inadequate attention has been given to modelling the possible relationship between the performance of a theory and the (monetary and nonmonetary) motivation of decision-makers.[8]

But other Duhem-Quine issues regularly arise. Both when the results are favorable and when they are unfavorable to a theory, experimental economists have asked if the observations were affected by increased subject experience. Thus, Alger (1986) reports oligopoly results in which early convergence to Nash behavior does not persist when much longer experiments are run. But Alger used simulated buyers, and it has been shown that mean prices are uniformly lower in oligopoly competition when real buyers are used (Brown-Kruse, 1991). These and a host of similar Duhem-Quine issues are subject to empirical examination and are part of the day-to-day operating life of experimentalists.

The "replication" problem is also related to Duhem-Quine. It is often claimed that there is inadequate replication in economics. The common complaint is that because replications are inadequately "original," editors are reluctant to publish them, and researchers are not well-motivated to conduct them. Experimental economists should perform replications, and often do so, as part of the process of reporting new experiments, so that the results can be compared with replications of previous studies. Of course, few such replications are completely pure: seldom does a researcher attempt to replicate exactly all the instructions, procedures, subject type, and other conditions used in a previous study. I would argue that such attempts at pure replication are in order only when the *results* of a previous study fail to replicate, and it is desirable to investigate why. If I do an experiment similar to yours as a baseline control for comparison with a related experiment I intend to perform, I am testing the robustness of your original results using my instructions, my subjects and a different experimenter. In effect, I am varying some of the more routine auxiliary hypotheses, and asking if the results are nonetheless indistinguishable. As a practical matter they most often are. When they are, then my experiment provides *more* support for the original theory than if the *same* (earlier) experiment was simply repeated. Franklin (1990, p. 107–8) makes this point by noting that if you want to know the correct time, it is more informative to compare your watch with another's than for either of you to look at your own watch twice. Intuitively, experimentalists and editors apply this principle in rejecting routine "pure replication" as not sufficiently original.

Experimentalists and other economists often use the rhetoric of "falsifying" theories, but it is clear from the totality of our professional conduct that falsification is just a means to a different end: the modification of theory in the light of evidence. To pursue this end, we need to know not only the conditions under which extant theory is falsified, but also the conditions under which it is verified. It is naive to suppose that any experiment will deliver the death blow to some theory. Theory always swims in the rough water of anomaly. You don't abandon a theory because of a (or many) falsifying observation(s). When Newton published the *Principia*, it was well-known that he could not even account for the orbit of the moon. Einstein's famous paper "On the Electrodynamics of Moving Bodies" (*Annalen der Physik*, 17, 1905) was "refuted" within a year by Kaufman (in the same journal) whose β-ray experiments showed that the deviations from the predictions of the theory were considerably beyond the limits of error that could be attributed to his equipment. Einstein agreed, but rationalized: "Only after a diverse body of observations becomes available will it be possible to decide with confidence whether systematic deviations are due to a *not yet recognized* source of errors or to the circumstance that the foundations of the theory of relativity do not correspond to the facts" (Einstein, 1907, p. 283, italics are mine). As it turned out, Kaufman's apparatus was later found to be faulty.

If you look at what experimental economists do, not what they say, you get the right picture of science learning. When a theory works well, they push imaginatively to find deliberately destructive experiments that will uncover its edges of validity, setting the stage for better theory and a better understanding of the phenomena. When a theory works poorly, they reexamine instructions for lack of clarity, increase the experience level of subjects, try increased payoffs, and explore sources of "error" in an attempt to find the limits of the falsifying conditions; again, this is for the purpose of better understanding the anatomy of a theory's failure, or the procedures for testing it, and thereby laying the basis for improving the theory. Ultimately, the procedures under which a theory is tested should be part of the theory.[9] But this step requires theorists' models to reflect a close understanding of the circumstances that produced the observations.

I am indebted to Timothy Taylor, Don McCloskey, and Alan Krueger for helpful comments and editing of an earlier version.

Notes

1. Where appropriate, references to work by me and my coauthors will be to the paper numbers in Smith (1991).
2. Selten (1989) offers a measure of predictive success. I use the terms "prediction" and "implication" of a theoretical model interchangeably. Consistency with a "prediction" does not require that the theory be done in advance of an observation.
3. That this description applies to markets in the field has been demonstrated by Forsythe et al. (1992), who report the remarkable forecasting accuracy of their presidential stock market, which beats the opinion polls by a wide margin.
4. Kachelmeier and Shehata (1992) report that these results also hold in cross-cultural comparisons of subjects from China, the United States and Canada.
5. Aumann (1987, p. 473) has emphasized in unmistakable terms this requirement of game theory: "It is not enough that each player be fully aware of the rules of the game and the utility functions of the players. Each player must also be aware of this fact . . . There is evidence that game theorists had been vaguely cognizant of the need for some such requirement ever since the late fifties or early sixties; but the first to give a clear, sharp formulation was the philosopher D. K. Lewis (in 1969). Lewis defined an event as common knowledge among a set of agents if all know it, all know that all know it, and so on ad infinitum. The common knowledge assumption underlies all of game theory and much of economic theory. Whatever

be the model under discussion, whether complete or incomplete information, consistent or inconsistent, repeated or one-shot, cooperative or noncooperative, the model itself must be assumed common knowledge; otherwise the model is insufficiently specified, and the analysis incoherent."

6. Of relevance here is the "getting to common knowledge" theorem discussed in this journal by Geanakoplos (1992). The theorem is driven by a process in which all agents observe in turn each agent's action. At some finite time, t^*, all agents have common knowledge of what each agent will do in the future. The asset experiments confirm the predictions of the theorem. But this does not imply that the subjects in the experiments go through a reasoning process like that which is used to prove the theorem. In fact, subjects would have great difficulty articulating the means whereby they reached their unwillingness to trade away from fundamental value.
7. At the other pole from Harrison stand some psychologists who downplay the evidence; that monetary payoffs can have a significant affect on outcomes. To wit: "We agree with Smith and Walker (1993) that monetary incentives could improve performance under certain conditions by eliminating careless errors" (Tversky and Kahneman, 1992, p. 316). The reader will not find any statement like this in the cited reference to agree with. The "errors" we discuss are not careless; they are deviations from optimality attributed to decision costs. If subjects care less about getting it right when there are zero or low rewards, and decision is costly, this is because it is in their interest to care less. We canvass 31 studies in which increasing rewards relative to baseline either reduces the deviations of the data around the theory's prediction, or moves the central tendency of the data closer to this prediction.
8. Of course, one can always offer the incredible argument that any recalcitrant case would go away if you just made payoff opportunity cost large enough. But this argument simply shows the limitations of a theory that postulates motivated agents, but is devoid of all detail as to that motivation. "Auxiliary" hypotheses in experimental economics that have to do with key issues involving the state of the agent like motivation and experience (learning), must ultimately be incorporated into the theory, not banished to the realm of auxiliary hypotheses for the experimentalists to worry about.
9. This is recognized by Bicchieri (1988), Brandenberger (1992), Geanakoplos (1992) and others when they model common knowledge as part of the theory of backward induction games. It is common for "background assumption" eventually to be made part of the theory.

References

Alger, Dan, *Investigating Oligopolies Within the Laboratory*, A Staff Report of the Bureau of Economics of the Federal Trade Commission, Washington, D.C., January 1986.

Aumann, Robert J., "Game Theory." In Eatwell, John, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics. 2.* London: MacMillian Press Ltd., 460–82.

Bicchieri, Cristina, "Backward Induction without Common Knowledge," *Philosophy of Science Association*, 1988, *2*, 329–43.

Bolton, Gary E., "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, December 1991, *81*:*5*, 1096–136.

Brandenberger, Adam, "Knowledge and Equilibrium in Games," *Journal of Economic Perspectives*, Fall 1992, *6*:4, 83–101.

Brown-Kruse, Jamie, "Contestability in the Presence of an Alternate Market: An Experimental Examination," *Rand Journal of Economics*, Spring 1991, *22*, 136–47.

Brown-Kruse, Jamie, "Laboratory Tests of Buyer Rationing Rules in Bertrand-Edgeworth Duopolies," working paper, *Department of Economics*, University of Colorado, 1992.

Camerer, Colin F., George Loewenstein, and Martin Weber, "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, October 1989, *97*:5, 1232–54.

Chu, Yun-Peng, and Ruey-Ling Chu, "The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note," *American Economic Review*, September 1990, *80*:4, 902–11.

Cox, James C., and Ronald Oaxaca, "Laboratory Experiments with a Finite-Horizon Job-Search Model," *Journal of Risk and Uncertainty*, September 1989, *2*:3, 301–29.

Cox, James C., and David Grether, "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives," working paper, *Economic Science Laboratory*, University of Arizona, July 1992.

Cox, James C., and Vernon L. Smith, "Endogenous Entry and Common Value Auctions," working paper, *Economic Science Laboratory*, University of Arizona, July 1992.

Davis, Douglas, and Charles A. Holt, *Experimental Economics*. Princeton: Princeton University Press, 1992.

Deng, Gang, et al., "Fairness: Effect on Temporary and Equilibrium Prices in Posted Offer Markets," working paper, *Department of Economics*, University of Arizona, October 1992.

Einstein, Albert, "On the Relativity Principle and the Conclusions Drawn from it," (1907). In *The Collected Papers of Albert Einstein* (translated by Anna Beck), volume 2. Princeton: Princeton University Press, [1989], 252–311.

Forsythe, Robert, Thomas Palfrey, and Charles Plott, "Asset Valuation in an Experimental Market," *Econometrica*, May 1982, *50*:3, 537–67.

Forsythe, R., J. Horowitz, N. Savin, and M. Sefton, "Replicability, Fairness and Pay in Experiments with Simple Bargaining Games," University of Iowa Working Paper 88–30, 1988, forthcoming in *Games and Economic Behavior*.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright, "Anatomy of an Experimental Political Stock Market," *American Economic Review*, December 1992, *82*:5, 1142–61.

Fouraker, Lawrence E., and Sidney Siegel, *Bargaining Behavior*. New York: McGraw-Hill, 1963.

Franciosi, Robert, Praveen Kujal, Roland Michelitsch, and Vernon L. Smith, "Experimental Tests of The Endowment Effect," working paper, *Economic Science Laboratory*, University of Arizona, March 1993.

Franklin, Allan, *Experiment, Right or Wrong*. Cambridge: Cambridge University Press, 1990.

Friedman, Daniel, and John Rust, eds., *The Double Auction Market: Institutions, Theories and Evidence*. Reading: Addison Wesley/SFI, 1992.

Friedman, Milton, *A Program for Monetary Stability*. New York: Fordham University Press, 1960.
Geanakoplos, John, "Common Knowledge," *Journal of Economic Perspectives*, Fall 1992, *6*:4, 53–82.
Gillette, David, and Robert DelMas, "Psycho-Economics: Studies in Decision Making," *Classroom Expernomics*, Newsletter published by Department of Economics, Management and Accounting, Marietta College, Fall 1992, *1*, 1–5.
Harrison, Glenn, "Theory and Misbehavior of First Price Auctions: Reply," *American Economic Review*, December 1992, *82*:5, 1426–43.
Hoffman, Elizabeth, "Bibliography of Experimental Economics," working paper, Department of Economics, University of Arizona, 1991.
Hoffman, Elizabeth, and Mathew Spitzer, "Entitlements, Rights and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice," *Journal of Legal Studies*, June 1985, *14*, 259–97.
Hoffman, Elizabeth, Kevin A. McCabe, Keith Shachat, and Vernon L. Smith, "Preferences, Property Rights and Anonymity in Bargaining Games," working paper, University of Arizona, September 1992, forthcoming in *Games and Economic Behavior*.
Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schadewald, "A Laboratory Market Examination of the Consumer Price Response to Information about Producer's Cost and Profits," *The Accounting Review*, October 1991, *66*, 694–717.
Kachelmeier, Steven J., and Mohamed Shehata, "Culture and Competition," *Journal of Economic Behavior and Organization*, October 1992, *19*, 145–68.
Kagel, John H., and Dan Levin, "The Winner's Curse and Public Information in Common Value Auctions," *American Economic Review*, December 1986, *76*:5, 894–920.
Kahneman, Daniel, Jack K. Knetsch, and Richard Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, September 1986, *76*:4, 728–41.
Kahneman, Daniel, Jack K. Knetsch, and Richard Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, December 1990, *98*, 1325–48.
Lichtenstein, Sarah, and Paul Slovic, "Reversals of Preference Between Bids and Choices in Gambling Decision," *Journal of Experimental Psychology*, July 1971, *89*, 46–55.
McCabe, Kevin A., "Fiat Money as a Store of Value in an Experimental Market," *Journal of Economic Behavior and Organization*, October 1989, *12*:2, 215–31.
McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith, "Designing a Uniform Price Double Auction." In Friedman, D., and J. Rust, eds., *The Double Auction Market; Institutions, Theories and Evidence*. Reading: Addison Wesley/SFI, 1992, 307–32.
Noussari, Charles, and David Porter, "Allocating Priority with Auctions," *Journal of Economic Behavior and Organization*, October 1992, *19*, 169–95.
Plott, Charles R., and Vernon L. Smith, "An Experimental Examination of Two Exchange Institutions," *Review of Economic Studies*, February 1978, *45*:1, 133–53.
Plott, Charles, "Some Policy Applications of Experimental Methods." In Roth, A. E., ed., *Laboratory Experimentation in Economics*. Cambridge: Cambridge University Press, 1987, 193–219.
Porter, David, and Vernon L. Smith, "Futures Markets, Dividend Certainty and Common Expectations in Asset Markets," mimeo, Department of Economics, University of Arizona, April 1989 (revised November 1992).

Roseveare, N. T., *Mercury's Perihelion from Le Verrier to Einstein.* Oxford, U.K.: Clarendon Press, 1982.

Roth, Alvin E., "Bargaining Phenomena and Bargaining Theory." In Roth, A. E., ed., *Laboratory Experimentation in Economics.* Cambridge, U.K.: Cambridge University Press, 1987, 14–41.

Samuelson, William, and Richard Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, March 1988, *1*:1, 7–59.

Schelling, T. C., "Bargaining, Communication and Limited War," *Journal of Conflict Resolution*, 1957, *1*, 19–36.

Selten, Reinhard, "Properties of a Measure of Predictive Success," University of Bonn Discussion Paper No. 13–130, October 1989.

Siegel, Sidney, and Lawrence E. Fouraker, *Bargaining and Group Decision Making: Experiments in Bilateral Monopoly.* New York: McGraw-Hill, 1960.

Simon, Herbert A., "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, February 1955, *69*, 99–118.

Smith, Vernon L., *Papers in Experimental Economics.* New York: Cambridge University Press, 1991.

Smith, Vernon L., and James M. Walker, "Monetary Rewards and Decision Cost in Experimental Economics," *Economic Inquiry*, April 1993, *31*, 245–61.

Thaler, Richard, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, March 1980, *1*:1, 39–60.

Tversky, Amos, and Daniel Kahneman, "Advances in Prospect Theory: Accumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, October 1992, *5*:4, 297–323.

Van Boening, Mark V., and Nathaniel T. Wilcox, "A Fixed Cost Exception to the Hayek Hypothesis." Paper presented at the Economic Science Association Meetings, Tucson, Arizona, October 24, 1992.

NINETEEN

Neuroeconomics

Using Neuroscience to Make Economic Predictions

Colin F. Camerer

Colin F. Camerer (1959–) was educated at Johns Hopkins and the University of Chicago and since 1994 has been a professor of economics at the California Institute of Technology. Camerer's research lies at the boundaries between cognitive psychology, neurophysiology, and economics. He is deeply involved in experimental economics, and his book, *Behavioral Game Theory*, is the most comprehensive recent survey of experimentation in economics.

Neuroeconomics seeks to ground economic theory in detailed neural mechanisms which are expressed mathematically and make behavioural predictions. One finding is that simple kinds of economising for life-and-death decisions (food, sex and danger) do occur in the brain as rational theories assume. Another set of findings appears to support the neural basis of constructs posited in behavioural economics, such as a preference for immediacy and nonlinear weighting of small and large probabilities. A third direction shows how understanding neural circuitry permits predictions and causal experiments which show state-dependence of revealed preference – except that states are biological and neural variables.

Neuroeconomics seeks to ground microeconomic theory in details about how the brain works (Zak, 2004; Camerer *et al.,* 2004; Chorvat and McCabe, 2005; Sanfey *et al.,* 2006). Neuroeconomics is a subfield of behavioural economics-behavioural economics which uses empirical evidence of limits on computation, willpower and greed to inspire new theories; see Mullainathan and Thaler, (2000); Camerer, (2005). It is also a subfield of

This article was prepared for the Hahn Lecture, Royal Economic Society, Nottingham UK, April 20, 2006. Thanks to all my collaborators whose joint work is reported (Ralph Adolphs, Meghana Bhatt, Ming Hsu, Michael Spezio, Dan Tranel, Joseph Wang), to RA's Min Rang and Alex Brown, to sceptics for forcing us to think harder and write more clearly about the enterprise, and to many neuroscientists (especially Ralph Adolf, John Allman, Paul Glimcher, John O'Doherty and Read Montague) for tutoring and advice over the last few years.
Economic Journal, vol. 117 (March, 2007): C26–C42. Reproduced by permission of Blackwell Publishing.

experimental economics because neuroeconomics requires mastery of difficult experimental tools which are new to economists (discussed in further detail in Section 1 below). And to many neuroscientists, the greatest promise of neuroeconomics is to supply theories and experimental designs for neuroscience. These neuroscientists feel that the kinds of models and tasks economists use routinely can contribute to 'systems neuroscience' understanding of higher-order cognition, which are challenging for neuroscientists who are used to focusing on very fine details of neurobiology and specific brain areas.

To modern economists, the neuroeconomic approach seems to be a sharp turn in economic thought. Around the turn of the nineteenth century, neoclassical economists made a clear methodological choice, to treat the mind as a black box and ignore its details for the purpose of economic theory (Bruni and Sugden, 2007). In an 1897 letter Pareto wrote

> It is an empirical fact that the natural sciences have progressed only when they have taken secondary principles as their point of departure, instead of trying to discover the essence of things . . . Pure political economy has therefore a great interest in relying as little as possible on the domain of psychology (quoted in Busino, 1964, p. xxiv).

Pareto's view that psychology should be deliberately ignored was partly reflective of a pessimism of his time, about the ability to ever understand the brain well enough to use neural detail as a basis for individual economising. (This pessimism was also manifested in the behaviourist psychology of Watson and Skinner, who turned attention away from the 'mentalism' of their time to stimulus-response relations and conditioning.)

As William Jevons wrote a little earlier, in '*Theory of Political Economy*'

> I hesitate to say that men will ever have the means of measuring directly the feelings of the human heart. It is from the quantitative effects of the feelings that we must estimate their comparative amounts (Jevons, 1871).

This turn-of-the-century pessimism about understanding the brain led directly to the rise of 'as if' rational choice models in neoclassical economics. Models of this sort posit individual behaviour which is consistent with logical principles, but do not put any evidentiary weight on direct tests of whether those principles are followed. For example, if a consumer's choices are transitive and complete, then she acts as if she attaches numerical utilities to bundles of goods and chooses the bundle with the highest utility, but direct measurement of utility is thought to be irrelevant as a test of the theory.

The ignorance of psychology that Pareto explicitly advocated was cemented by Milton Friedman's (1953) development of 'positive economics'. Friedman, and the many economists influenced by his view, advocated two principles for judging theories which use assumptions *A* to make a formal prediction *P*:

1. Assumptions *A* should be judged by the accuracy of the predictions *P* they mathematically imply.
2. Since false assumptions can yield accurate predictions, even if assumptions appear false their empirical weakness should be tolerated if they lead to accurate predictions *P*.

I wholeheartedly endorse the first principle (1), but not the corollary principle (2).

Here is why: first, if assumptions *A* are false but lead to an accurate prediction, they presumably do so because of a hidden 'repair' condition *R* (that is, (not-A and *R*) $\rightarrow$ P is a more complete theory at both ends than $A \rightarrow P$). Then the proper focus of progressive research should be specifying the repair assumption *R* and exploring its implications, in conjunction with more accurate assumptions.

Second, the importance of making good predictions (1) is precisely the reason to explore alternative assumptions grounded in psychological and neuroscientific facts. We do this in behavioural economics because we hope that models based on more accurate assumptions will make some interesting new predictions, and better predictions overall.

As-if models based on dubious assumptions clearly work well in many respects, and always will (just as expected value is still a useful tool for some kinds of analysis, even though it is a severe restriction of expected utility). But tests of the predictions that follow from as-if rational choice have also established many empirical anomalies. Behavioural economics describes these regularities and suggests formal models to explain them (Camerer, 2007).

Debates between rational-choice and behavioural models usually revolve around psychological constructs, such as loss-aversion (Kahneman and Tversky, 1979), the role of learning and limited strategic thinking, a preference for immediate rewards, and precise preferences over social allocations, which have not been observed directly. But technology now allows us to open the black box of the mind and observe brain activity directly. These direct observations can only enhance the development of theories which are based on more accurate assumptions *and* make better predictions as a result.

An analogy to organisational economics illustrates the potential of neuroeconomics (Sanfey *et al.*, 2006). Until the 1970s, the 'theory of the firm' was basically a reduced-form model of how capital and labour are combined to create a production function. The idea that a firm just combines labour and capital is obviously a gross simplification – it neglects the details of principal-agent relations, gift exchange and efficiency wages, social networks and favour exchange in firms, substitution of authority for pricing, corporate culture and so forth. But the gross simplification is useful, for the purpose of building up an industry supply curve.

Later, contract theory opened up the black-box of the firm and modelled the details of the nexus of contracts between shareholders, workers and managers. The *new* theory of the *firm* replaces the (perennially useful) fiction *of a profit-maximising firm* which has a single goal, with a more detailed account of how *components* of the *firm – individuals, hierarchies,* and *networks* – interact and communicate to determine *firm* behaviour.

Neuroeconomics proposes to do the same by treating an individual economic agent like a firm. The last sentence in the previous paragraph can be exactly rewritten to replace firms and the components of firms with individuals and neural components of individuals. Rewriting that sentence gives this one: The *neuroeconomic* theory of the *individual* replaces the (perennially useful) fiction of a *utility-maximising individual* which has a single goal, with a more detailed account of how *components* of the *individual – brain regions, cognitive control,* and *neural circuits* – interact and communicate to determine *individual* behaviour.

The rapid emergence of various dual-self or dual-process approaches testifies to how well economic theory can be adapted to study the brain as an organisation of interacting components. Fudenberg and Levine (forthcoming) emphasise the struggle between a long-run player and a short-run player, adapted from game-theoretic models (see also Shefrin and Thaler's prescient, 1988, 'planner-doer' model).[1] Benhabib and Bisin (2005) emphasise the constraint that controlled 'executive' processes put on automatic processes. Bernheim and Rangel (forthcoming) emphasise 'hot' impulsive states (akin to automatic process, but perhaps driven by visceral factors like drug craving or hunger) and 'cold' states. Loewenstein and O'Donoghue (2004) emphasise deliberate processes and affective ones. Brocas and Carillo (2005) emphasise how a cortical control process constrains an emotional process which may be asymmetrically informed. So far, there is little direct neural evidence testing these various models and comparing them. Doing so is an obvious immediate direction for research (and will contribute to basic neuroscience as well).

It is important to note that the focus of neuroeconomic research so far is largely on microeconomics foundations of consumer choice, valuing risky gambles, and strategic thinking. It remains to be seen whether neural measurement will be useful for understanding macroeconomic phenomena like consumer confidence or stock market bubbles. However, many of these macro phenomena might spring from the interaction of many brains that are tightly linked through social networks and common responses to emotional and news shocks which can be reciprocal or contagious. If so, macro models could explore how the result of brain activity has a multiplier effect in the economy.

1. Neuroscientific Facts and Tools

1.1. Facts

Some basic facts about the human brain are useful for economists to know, to understand the evidence presented below and to provide constraint on theorising.

The brain is weakly modular, in the sense that not every brain area contributes to every behaviour. (That is, the early phrenologists were on the right track, but had too crude a concept of how localised complex behaviour or traits like 'virtue' and 'sloth' were.) While the brain is modular, it is also 'plastic' – responsive to environment as brain 'software' is gradually 'installed'. Plasticity is most obvious in childhood development but seems to continue well into adolescence. Plasticity is the reason why neuroscientists usually bristle at the term 'hard-wired', which economists often use casually.

While neuroscientists often focus on specific brain areas which are cytoarchitecturally distinct (i.e., they have distinct tissue, neurons, and neurotransmitters), for tasks economists are interested in the proper focus is 'circuits' of multiple brain areas. The importance of circuitry also implies that the right kinds of models are computational ones in which well-understood components collaborate to create behaviour.

Attention and consciousness are scarce, and the brain is evolved to offload decisions by automating activity through learning. Automaticity means that people are capable of creating tremendous expertise which relies on subconscious intuition and pattern recognition. It also means that overcoming automated behaviour takes scarce conscious effort and is often a source of mistakes in 'Stroop tasks'.[2]

The human brain is basically the primate brain with extra neocortex; and the primate brain is a simpler mammalian brain with some neocortex. This evolutionary history is the main reason why experiments with animals are so informative about human behaviour. (To think otherwise is economic creationism.) For example, rats become biologically addicted to all substances that humans become addicted to (nicotine, opiates, alcohol etc.). Our shared evolutionary past, and inherited brain regions, do not imply that humans always behave like monkeys (though we sometimes do). Our shared past just implies that when humans struggle to control animal impulses (such as drug addiction), the struggle is between the neocortex and older temporal-lobe areas. Knowing which areas are involved in the struggle is useful for crafting theory and for prescribing treatments.

1.2. Tools

Much of the potential of neuroeconomics comes from relatively recent improvements in technology for measuring brain activity (particularly fMRI), and in matching older technologies (such as eyetracking and EEG) with new tasks.

fMRI uses magnetic resonance imaging, popular for decades for medical diagnosis, at rapid frequencies to measure oxygenated blood flow in the brain (which is correlated with neural input). The spatial resolution of fMRI is about 3 cubic millimetre voxels and its temporal resolution is 2 seconds. Stronger magnetic fields are unlikely to provide much more improvement (and may pose health risks, which modern 3-tesla magnets do not); but improvement may come from innovation in experimental design and statistics.

Positron Emission Tomography (PET) is an earlier scanning technology which injects radioactive solution (usually glucose with a radioactive marker). PET temporal resolution is worse than fMRI (minutes rather than seconds) but glucose is a more direct correlate of neural activity than blood flow.

fMRI and PET are good for roughly identifying areas that are active in a task. Once candidate circuits are established, it is useful to ask whether behaviour is changed when parts of the circuit are broken or disrupted.

Studies of patients with brain lesions are useful for testing hypotheses from fMRI. If a patient with damage to area X cannot perform a task T normally, then area X is part of a normal circuit for doing T. (Lesion data are reported below in a study of the Ellsberg paradox in ambiguous choice.) Transcranial magnetic stimulation (TMS) can 'knock out' or activate brain areas, and hence is useful for knowing what targeted areas do. The animal model is also useful because invasive surgeries and genetic engineering can be done with animals, as a substitute for exogenous lesions and correlational studies in humans.

A much more detailed level of data comes from recording activity of a single neuron at a time, mostly from primates (and, rarely, from human neurosurgical patients in whom electrodes have been planted to detect locations of epileptic seizures to locate surgical targets).

Older tools continue to be useful. The electroencephalogram (EEG) records very rapid (millisecond) electrical activity from outer brain areas, and can sometimes be used to interpolate activity in areas deeper in the brain. Psychophysiological recording (of skin conductance, heart rate and pupil dilation, for example) are cheap and easy too. Tracking where people are looking on a screen (eyetracking) is also very easy and useful for many questions economists ask. Directly observing the information people use to make decisions provides a second dependent variable that can be used, in conjunction with observed choices, to identify decision rules better than choices alone can.

A great strength of neuroscience is that investigators who have mastered these tools compete fiercely (for grants, students, and space in *Science* and *Nature);* their fierce competition creates a bonus for methodological innovation and weeds out weak results. The tools are also complements because each tool can compensate for the weaknesses of others (e.g., having an fMRI finding makes data from patients with lesions in the areas identified by fMRI especially valuable). Recognising this complementarity, neuro-scientists are most comfortable with ideas that are consistent with many types of data recorded in different ways at different levels of temporal and spatial resolution. Happily for economists, many of our simplest questions can be illuminated by the simple measures (e.g., eye tracking and psychophysiological recording). Ambitious graduate students interested in this field are well advised to pick one tool that can help answer the questions they are interested in, and master it.

Neuroeconomics is likely to provide three types of evidence about economic behaviour. Examples of each type of evidence are given in the next three Sections of this paper.[3] The three kinds of evidence are:

1. Evidence which shows mechanisms that implement rational choice (utility-maximisation and Bayesian integration of information), typically in tasks that are highly-sculpted to make decisions that are useful for survival across species (vision, food, sex and danger).
2. Evidence which supports the kinds of variables and parameters introduced in behavioural economics.
3. Evidence which suggests the influence of 'new' variables that are implicit, underweighted, or missing in rational-choice theory.

2. Evidence for Rational Choice Principles

In many simple choice domains, evolution has had a long time to sculpt cross-species mechanisms that are crucial for survival and reproduction (involving food, sex and danger). In these domains, evolution has either created neural circuits which approximate Bayesian-rational choice, or learning mechanisms that generate Bayesian-rational choice with sufficient experience in a stationary environment, putting to use highly-developed capacities for sensory evaluation (vision, taste, smell, sound), memory and social imitation.

For example, Platt and Glimcher (1999) find remarkable neurons in monkey lateral intraparietal cortex (LIP) which fire at a rate that is almost perfectly correlated with the expected value of an upcoming juice reward, triggered by a monkey eye movement (saccade); see also Bayer and Glimcher (2005). Deaner *et al.* (2005) find that monkeys can reliably trade off juice rewards with exposure to visual images (including images of females from behind and faces of high and low status conspecific monkeys). Monkeys can also learn to approximate mixed-strategies in games (Glimcher *et al.*, 2005), probably using generalised reinforcement algorithms (Lee *et al.*, 2004). Neuroscientists are also finding prefrontal neurons that appear to express values of choices (Padoa-Schioppa and Assad, 2006) and potential locations of 'neural currency' that creates tradeoffs (Conover and Shizgal, 2005). Following a long tradition in 'animal economies' (Kagel *et al.*, 1995), Chen *et al.* (2006) show that capuchin monkeys respond to price changes, obeying the GARP axiom, when exchanging tokens for different food rewards.

Another literature shows that Bayesian models are accurate approximations of how different kinds of sensory information are integrated (Stocker and Simoncelli, 2006). These data are in sharp contrast with many cognitive psychology experiments showing that Bayesian principles are violated when intelligent humans evaluate abstract events (Kahneman, 2003). It is difficult to reconcile these two literatures directly, because it is difficult to create tasks in which monkeys have to judge the kind of abstract questions people are asked – like whether basketball players have a 'hot hand' or whether representative conjunctions of events (F & B) are more likely than their component events (F and B judged separately). Common paradigms that can be used across species represent a huge challenge that would be very useful for either reconciling the results across species or establishing why they differ.

3. Evidence for Behavioural Economics Principles

This Section discusses four areas in which neuroscience has established some tentative neural foundation for ideas from behavioural economics which were derived earlier from experiments and field data. The four areas are: $\beta - \delta$ time discounting; aversion to missing information about probability (ambiguity); nonlinear weighting of probability; and limited strategic thinking in games.

3.1. Time Discounting

Extensive experiments with animals, and later with humans, established that the discount factor put on future rewards is closer to a hyperbola, $1/(1+kt)$, than an exponentially-declining discount factor δ^t. Laibson (1997) borrowed a two-piece discounting function introduced to explain parental bequests, to model 'quasi-hyperbolic' discounting. In the $\beta-\delta$ model, agents put a weight of one on current rewards, and weight future rewards at discrete time $t>0$ by $\beta\delta^t$. (When $\beta=1$ the two-parameter function reduces to an exponential.) O'Donoghue and Rabin (1999) dubbed the β term a 'present bias' and explore its implications. Various field and experimental data suggest values of β around 0.6–0.8.[4] To search for β and δ processes in the brain, McClure *et al.* (2004) presented subjects with choices between a current reward and a reward with a one-month delay (which activates both β and δ systems), and other choices with a one-month or two-month delay (in which the β component divides out and only δ remains). They find activity in areas often associated with an emotional limbic system (medial frontal cortex, cingulate and ventral striatum) when β comes into play, and find distinct activity in lateral orbitofrontal cortex and dorsolateral cortex linked to the δ system. Their study is hardly the last word – in fact, it is the first word – but is consistent with discounting being a splice of two processes.

3.2. Ambiguity-Aversion

In subjective expected utility theory, the willingness to take bets on events is taken to reveal subjective probabilities of those events. The Ellsberg paradox showed that for a small majority of subjects, when two events are equally likely but poorly understood (or 'ambiguous'), revealed decision weights seem to combine judgment of likelihood and an additional factor which leads to an aversion to betting under ambiguity. Theories of nonadditive probability and set-valued probabilities loosely ascribe this ambiguity-aversion to pessimism or fear of betting in the face of unknown information. Ambiguity-aversion has been implicated in 'home bias' in financial investment (a preference for investing in stocks in one's own country, or firm, or firms nearby), in 'robust control' in macroeconomics, and in other economic domains (Hsu *et al.*, 2005). Scottish law provides a useful practical example.

In Scottish law there are three verdicts – guilty, not guilty and 'unproven'. An unproven verdict results when there is too little evidence to determine guilt or innocence (often in sexual assault cases, since Scottish law requires a corroborating witness besides a testifying victim). Unproven verdicts are usually the jury's way of expressing an aversion to rendering either verdict, often shaming a victim they believe is guilty but cannot legally find guilty because of evidentiary rules which create reasonable doubt.

Since decision theorists forming axioms are not generally thinking about brain activity adhering to those axioms, it is difficult to find descriptions which are suggestive of neural activity. But Raiffa (1961) wrote:

> But if certain uncertainties in the problem were in cloudy or fuzzy [ambiguous] form, then very often there was a shifting of gears and no effort at all was made to think deliberately and reflectively about the problem. Systematic decomposition of the problem was shunned and an over-all 'seat of the pants' judgment was made which graphically reflected the temperament of the decision maker.

Unfortunately, the 'seat of the pants' is not a brain area, but Raiffa does describe a rapid emotional response in the face of ambiguity. Hsu *et al.* (2005) investigated ambiguity and risk using fMRI; see also Huettel *et al.* (2006). They found additional activation in valuing bets on ambiguous gambles relative to risky ones (such as bets on low-knowledge events, like the temperature in Tajikistan compared to high-knowledge New York). They found additional activity in the dorsolateral prefrontal area, orbito-frontal cortex (above the eye sockets, OFC) and the amygdala (a 'vigilance' area, which is rapidly activated in 5–20 msecs by fearful images, even before they are consciously processed). Subjects with higher right OFC activity in response to ambiguity also had higher ambiguity-aversion parameters as estimated by a stochastic choice logit model fit to gamble valuations.

3.3. Nonlinear Probability Weighting

In expected utility (EU) theory, the utilities of gamble outcomes are weighted by their probability p. But many experimental studies suggest that people actually weight probabilities nonlinearly with a function $\pi(p)$, overweighting low probabilities and underweighting probabilities close to one (the 'certainty effect'); see Prelec (1998). Overweighting of low p could be important in pricing insurance and in explaining demand for lottery tickets and the high failure rate of new businesses.

Measuring neural activation in response to variation in probability is made possible by the fact that a fair amount is known about how the caudate (a temporal lobe area including the striatum) responds to anticipated reward. Hsu *et al.* (2006) set out to see whether activation in the striatum responded nonlinearly to probability of winning. They first presented simple binary gambles (p, *X*) which have a p chance of paying *$X* (otherwise they pay zero) for a few seconds, then had subjects choose between the presented gamble and a second gamble (roughly matched for expected value). The choice data enable estimation of parameters of a probability weighting function $\pi(p)$. They look at activity in the left and right caudate areas – an area in the temporal lobe associated with rewards of many types (juice, cocaine, attractive faces, money, faces of people who have cooperated with you). Controlling for the payoff amount *X*, there is a modest nonlinearity of activity across levels of probability p which is reasonably similar to the nonlinear functions shown in Prelec (1998). This similarity of indirect estimates and direct estimates of caudate activity is not conclusive proof that the brain is weighing probabilities nonlinearly, but it is consistent with that hypothesis. A likely explanation is that probability estimation is a combination of a linear weighting and an inverse-S step function which sorts probabilities crudely into 'no, maybe, yes'.[5] Combining the two gives a regressive function that overweighs low p and underweighs high p, and is consistent with the brain activation.

3.4. Limited Strategic Thinking

In game theory, players are in equilibrium when they optimise and guess correctly what other players will do – that is, when their beliefs about other players' strategies match the actual strategies others choose. Camerer *et al.* (2004) describe an alternative 'cognitive hierarchy' (CH) theory in which players use various steps of strategic thinking. Some step-0 players randomise, other step-1 players anticipate randomisation and best-respond to it, step-2 players best-respond to a mixture of step-0 and step-1 players, and so on. Since the highest-step players anticipate correctly the distribution of what other players will do, their beliefs are in equilibrium, but the beliefs of lower-step thinkers are not in equilibrium because they do not guess correctly what higher-step players will do. This model (and earlier versions introduced by others) fits empirical data from dozens of game experiments with many different structural forms (mixed-equilibria, coordination, dominance-solvable games and so forth).

To look for evidence of limited strategic thinking in the brain, Bhatt and Camerer (2005) did fMRI of players when they made choices, and when they expressed beliefs about what other players would do. They found that

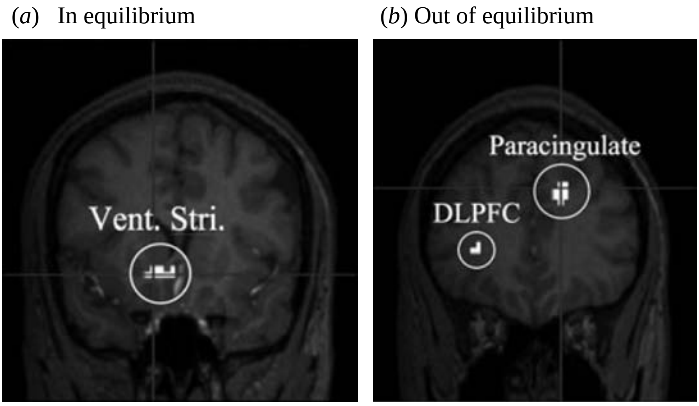

Figure 19.1. Differences in Brain Activity During Choosing a Strategy and Expressing a Belief About Another Player's Strategy (Bhatt and Camerer, 2005).

Equilibrium trials (a) show only a difference in ventral striatum (a Reward Anticipation Area). Out-of-equilibrium trials (b) show stronger activity in choosing than in belief expression (highlighting paracingulate and dorsolateral prefrontal (DLPFC) areas), which suggests subjects are not reasoning strategically about other players.

when players' choices and beliefs were in equilibrium, there was almost perfect overlap in brain activity during choosing and belief expression – that is, creating equilibrium beliefs requires players to imagine how others are choosing, which uses overlapping neural circuitry with making your own choice (Figure 19.1). When players were out of equilibrium, there was much more activity when making a choice than when expressing a belief (as would be expected from 0 and 1-step thinkers, who are thinking harder about their own choice than they are about choices of other players). Thus, being in equilibrium is not merely a mathematical restriction on equality of choices and beliefs, it is also a 'state of mind' identifiable by brain imaging.

4. Evidence for New Psychological Variables

The largest payoff from neuroeconomics will not come from finding rational-choice processes in the brain for complex economic decisions, or from supporting ideas in behavioural economics derived from experimental and field data (as shown by examples in the last two sections). The largest innovation may come from pointing to biological variables which have a large influence on behaviour and are underweighted or ignored in standard theory. This section lists a few speculative examples. They suggest that the concept of a preference is not a primitive (as Pareto suggested); preferences are both the output of a neural choice process, and an input which can be used in economic theory to study responses to changes in prices and wealth. This view implies that if we understand what variables affect preferences, we can shift preferences and shift behaviour (without changing prices or constraints). Whether this can be done reliably or on a large scale is not yet known. The goal at this point is just to show that understanding biology and the brain can make fresh predictions about observed choices. At this point, there are few such predictions and they focus on small effects at the individual level. But given the youth of the field, having any such examples is suggestive and they point in interesting directions.

1. In the ambiguity study described in the last Section (Hsu *et al.*, 2005), there is a modest correlation of right OFC activity with a parameter characterising the degree of ambiguity-aversion, which is derived from estimation using choices. (The parameter γ is derived implicitly from the weight $(E(p)^{\gamma})$ given to an event with expected or diffuse-prior probability p. The value $\gamma = 1$ is ambiguity-neutrality. A value $\gamma > 1$ corresponds to ambiguity-aversion; an ambiguity-averse person acts as if the decision weight on an ambiguous event is lower than its expected probability.) One can extrapolate statistically from the correlation between OFC activation and γ in normal subjects to infer the behavioural value of γ that would be revealed by choices of a person with no OFC activity at all – due to a lesion in that area, say (see Figure 19.2). The extrapolated estimate is $\gamma = 0.85$ (roughly ambiguity-neutral, given sampling error). In fact, Hsu *et al.* also tested Ellsberg-type problems on patients with OFC damage subsuming the areas observed in fMRI. Those patients' choices exhibited a value of $\gamma = 0.82$. I would love to say this value was truly predicted before the fact, but it was not (both studies were conducted in parallel). In any case, there is a close link between the behavioural parameter 'predicted' by extrapolating from the fMRI evidence to patients with no activity, and the extrapolated parameter is close to the figure revealed by choices. While this correspondence could be construed as consistent with axiomatic theories of ambiguity-aversion, no theory would have predicted it without the fMRI evidence to tell us what lesion patients would be roughly ambiguity-neutral.
2. Wang *et al.* (2006) studied experimentally a classic 'biased-transmission game' that has been widely used in economics and

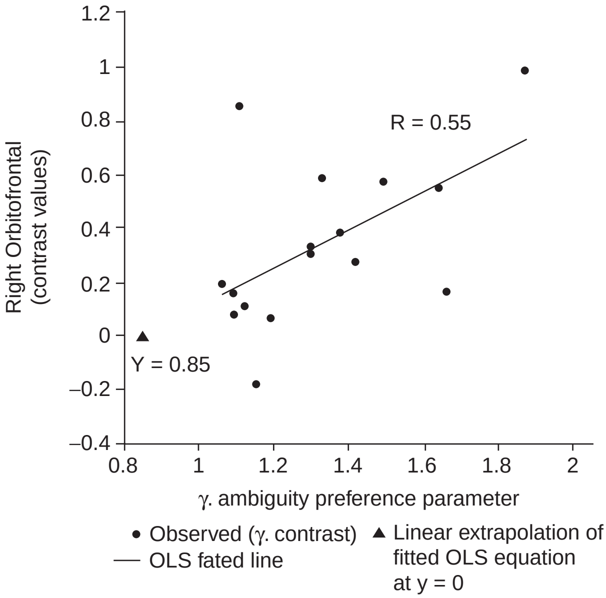

Figure 19.2. Correlation Between Individual-specific Ambiguity-aversion Parameters γ Estimated from Choices (x-axis, Higher γ is more Ambiguity-aversion) and Differential Activity in Right Orbitofrontal cortex in Ambiguous vs. Risky Gamble Evaluation (y-axis).

Positive correlation ($r = 0.55$) indicates more ambiguity-averse people have more differential activity in ROFC. Extrapolating to a person with no OFC activity ($y = 0$) gives an inferred ambiguity-aversion γ of 0.85. The actual behavioural parameter derived from choices of patients with OFC lesions was $\gamma = 0.82$.

political science. In this game, a sender observes a state *S*, an integer from 1 to 5 (uniformly distributed). The sender then chooses an integer message *M* from 1 to 5. A receiver knows the setup of the game, and learns the message, but *does not know the true state directly.* The receiver then chooses an action *A* from 1 to 5. (The game is like security analysts who know more about the value of a stock than you do, make a recommendation, and want you to act as if the stock is more valuable than it is, because of career concerns or other collateral interests.) In the interesting conditions, the senders earn the most if the receiver chooses $S + b$, where b is a known bias parameter (either 1 or 2). We try to predict the true state from the sender's message *M*, and from their pupil dilation (expansion of pupils) when they send their message. Pupils dilate under arousal, stress and deception (that is why poker players wear sunglasses if they are allowed to). Statistical tests show that measuring the pupil dilation improves substantially in predicting what the true state is. Thus, a biological variable helps infer private information which is conveyed by messages, in a way that is not explicitly predicted by conventional game theory.

3. Sanfey *et al.* (2003) used fMRI to see what areas were differentially active in the brains of responders in an ultimatum game, when the responders received a fair offer (\$4 − 5 out of \$10) compared to an unfair offer (\$1 − 2). They found activation in the insula (a discomfort or disgust area, perhaps measuring the emotional reaction to getting a low offer), dorsolateral prefrontal cortex (DLPFC, a planning and evaluation area), and anterior cingulate (a conflict-resolution area). They also found that whether people rejected low offers or not could be predicted with some accuracy from whether the insula was more active than DLPFC or *vice versa*. Building on this study, Wout *et al.* (2005) and Knoch *et al.* (2006) used repetitive TMS to disrupt the DLPFC when people received offers. Based on the fMRI evidence, they hypothesised that if the DLPFC is disrupted, the socialised response to unfairness which leads to rejection may be turned off, so that people will exhibit more innate selfishness and accept lower offers more often. Their prediction was correct. The effects are small and come from only two studies with modest sample sizes, but they show the power of a two-step process: first establish parts of neural circuitry that implement a behaviour; then stimulate or disrupt some of those parts and if see if you can *cause* a behavioural change.
4. Oxytocin is a powerful hormone in social bonding (e.g., it surges when mothers breast-feed; and synthetic oxytocin – pitocin – is administered in American hospitals to stimulate childbirth). Direct measurement from blood samples (Zak *et al.*, 2005) suggests oxytocin is important in trust. Inspired by this evidence, Kosfeld *et al.* (2005) had subjects play a trust game in which one player could choose whether to invest money or keep it. If she invested, the money doubled in amount and the responder player (the trustee) could decide how much to repay and how much to keep. Half the subjects were given a synthetic oxytocin dose (three puffs in each nostril, then wait an hour) and half were given a placebo so the subjects could not tell whether they got the real pitocin or nothing. Kosfeld *et al.* hypothesised that oxytocin would increase trust, and it did. Game theory makes predictions about structural variables that might increase trust – most reliably, whether the game is repeated or played once (which does have a strong impact; e.g., Chong *et al.* (forthcoming). But nothing in game theory would have predicted the effect of synthetic oxytocin.

5. Conclusion

The goal of neuroeconomics is to ground economic theory in details of how the brain works in decision making, strategic thinking, and exchange. One way to achieve this is to observe processes and constructs which are typically considered unobservable, to decide between many theories of behavioural anomalies like risk aversion, altruistic punishment, and reciprocity.

I have presented examples in which neuroeconomic evidence points to any of three conclusions. Sometimes rational-choice processes are clearly evident in brain activity (LIP neurons that fire at rates almost exactly linear in expected reward). In other cases, the variables or differences predicted by behavioural economics models are evident – in $\beta - \delta$ discounting, ambiguity-aversion and nonlinear probability weighting. In still another case, perhaps the most innovative, variables that are not a traditional focus of economic theory have perceptible effects and, sometimes, strong effects: patients with OFC damage are unusually ambiguity-neutral (which is consistent with fMRI evidence identifying the OFC as a locus of ambiguity-aversion processing); pupil dilation helps predict a player's private information when they might be lying; stimulating DLPFC increases acceptance of low ultimatum offers (because earlier fMRI work showed DLPFC activity is correlated with acceptance); and administering oxytocin makes people more trusting.

Thinking about how the brain implements economic decisions, compared to thinking about choices resulting from preference and belief, is like switching from watching TV in black and white to watching in colour – there are so many more variables to think about. For economic theorists, a natural way to think about these phenomena is that many biological state variables influence preferences; given those state-dependent preferences, prices and budget constraints have familiar influences. I agree with this view, except that we will never fully understand the nature of the state-dependence without facts from psychology and neuroscience. Furthermore, it is not clear whether subjects are aware of exogenous influences that alter these internal states and how the state-dependence works when a lot of money is on the line (arousal itself can be a big state variable) and when agents are highly experienced.

There is much obvious future research. One path is to study the multiple-process approaches seriously and look for those processes directly in the brain, or as they are manifested in behavioural experiments.[6] Another is to search for evidence of distinctions that are well-established in behavioural economics (such as gain-loss differences, framing effects, emotional foundations of inequality-aversion or social image, and so forth). A more unifying approach is to take the revealed-preference model seriously and see how far its language can be stretched to accommodate neural evidence, while making new predictions rather than just giving economic names to neural processes.

6. Afterword and Prologue: The 'Mindless' Critique, and a Reply from the Past

Some economists feel that the central theory in economics – revelation of inherently unobservable preferences and beliefs by observed choices – is immune to empirical evidence from neuroeconomics. Their argument is that economics is only about explaining choices, and neural evidence is not choices. For example, Gul and Pesendorfer (2005) suggest one categorisation of economics (which could be called 'economics™', because they so sharply legislate what economics is and is not). They write[7]

> ...the requirement that economic™ theories simultaneously account for economic™ data and brain imaging data places an unreasonable burden on economic™ theories (Gul and Pesendorfer, 2005)

Some of the examples in Sections 3 and 4 were judiciously chosen to address precisely this critique. Theories of $\beta - \delta$ time discounting and non-linear $\pi(p)$ probability weighting can account for both behavioural data from many choice experiments (and many field data too) *and* are consistent with tentative evidence of neural activity. Since such theories are possible, is it really an 'unreasonable burden' to ask whether other theories can do the same? Of course, theories that spring from the fertile mind of a theorist who is simply inspired by psychology, but is not beholden to a large body of facts, could prove to be useful theories too. But theories that can explain neural facts *and* choices should have some advantage over theories which explain *only* choices, if they are comparably tractable.

More fundamentally, the argument against neuroeconomics (or the case for 'mindless' economics, as their paper's title calls it) rests mostly on an interesting hope, and rests a little bit on the history of economic thought. The hope is that all anomalies produced by behavioural economics and neuroeconomics can be explained (if not predicted) by the enriched language of economics – preferences, beliefs, and imperfect information and constraint. I share that hope, but only if some imperfections and constraints are allowed to be located in the brain – in which case, brain evidence is useful for understanding those imperfections and constraints and suggesting the best models of them.

A useful focus for debate is therefore how gracefully (and predictively) conventional economics language can explain the effects on observed choices (and inferred unobservable states) of brain lesions, pupil dilation, TMS stimulation, and oxytocin. Any conventional accounts which absorb these effects semantically, and then make predictions about them, will be welcomed as interesting neuroeconomics.

The history of economic thought part of the 'mindless' case is more clearly settled. Gul and Pesendorfer write that 'Populating economic™ models with 'flesh-and-blood human beings' was never the objective of economists™'. But Colander (2005) reminds us how interested classical economists were in measuring concepts like utility directly, before Pareto and the neoclassicals gave up.

Edgeworth dreamed of a 'hedonimeter' that could measure utility directly; Ramsey fantasised about a 'psychogalvanometer'; and Irving Fisher wrote extensively, and with a time lag due to frustration, about how utility could be measured directly. Edgeworth wrote:

> ... imagine an ideally perfect instrument, a psychophysical machine, continually registering the height of pleasure experienced by an individual ... From moment to moment the hedonimeter varies; the delicate index now flickering with the flutter of the passions, now steadied by intellectual activity, low sunk whole hours in the neighbourhood of zero, or momentarily springing up towards infinity ...

The interest of these early economists in measuring utility directly was to establish a biological cardinal utility scale, which is not a goal of microeconomics. In any case, given their ambitions, it is hard to believe at least some of these important figures would not be interested in using the modern tools that we do have. If Edgeworth were alive today, would he just be making boxes, or also recording the brain?

Notes

1. Benabou and Pyciak (2002) show how the Gul and Pesendorfer (2001) model of preferences under temptation is mathematically equivalent to a rent-seeking competition between two brain areas, linking the preferential approach to the multiple-selves approach.
2. In the classic Stroop task, people are asked to name the colour of ink a word is printed in. Under time pressure, people invariably state the word rather than the colour (e.g., if the word 'black' is printed in green ink, they say 'black', not 'green') at first, though they can learn over time. The Stroop task is now used as a generic term for any automated response which must be overridden by cognitive control. The game 'Simon says' is an example. Another example is when Americans visit England. Americans are used to looking to the left for cars approach them when they cross the street but in England cars approach from the

right. Many Americans are killed every year because of a Stroop mistake. The fact that avoiding a Stroop mistake takes conscious effort also predicts that Americans whose conscious attention is absorbed elsewhere when they are crossing the street in England – talking on a cell phone, for example – are more likely to be killed than those who are not distracted.

3. Note that the length of the three sections is not intended to reflect either the accumulated regularity in each of the three areas, or likely future results. The last Section is longer because it presents a more novel perspective, and most directly meets the critique that neuroeconomics does not provide new insight.
4. Angeletos *et al.* (2001), Delia Vigna and Paserman (2005), Tanaka *et al.* (2006, http://www.hss. caltech.edu/∼camerer/Growth-nth.pdf) and Brown *et al.* (2006) all report estimates from savings data, unemployment data, abstract experiments in Vietnam, and dynamic savings rewards with temptation (respectively) with β around 0.6–0.8.
5. Attention and adaptation probably also play crucial roles. While some risks are overweighed, others might be dismissed entirely because they are not imagined or attended to. There is no experimental paradigm to turn on and off attention to low probability risks; having one would be useful, as would field measurements of actual attention to risks.
6. For example, the Bernheim-Rangel, Fudenberg-Levine, and $\beta - \delta$ time preference models all predict that subjects who are tempted by immediate rewards will make different decisions if current choices are not consumed until a time sufficiently far in the future (so that the 'hot self', 'short-run player', or 'present-biased' current player's myopic preferences are disabled). Brown *et al.* (2006) find the first direct evidence of such an effect in dynamic savings experiments, when thirsty subjects decide how much of a thirst-slaking beverage to consume. When subjects have to 'order in advance', by making choices at period t which are not consumed until period $t + 10$, they consume less and earn more overall rewards. Calibrating $\beta - \delta$ parameters to actual decisions yields sensible estimates of $\delta = 0.90$ and $\beta = 0.62-0.72$ (the latter depends on whether agents are sophisticated about their present bias, or naïve).
7. In the passages quoted from their paper, of course, the TM superscripts do not appear.

References

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J. and Weinberg, S. (2001). The hyperbolic consumption model: calibration, simulation, and empirical evaluation', *Journal of Economic Perspectives,* vol. 15 (3) (Summer), pp. 47–68.

Bayer, H. M. and Glimcher, P. W. (2005). 'Midbrain dopamine neurons encode a quantitative reward prediction error signal', *Neuron,* vol. 47 (1) (07 July), pp. 129–41.

Benabou, R. and Pyciak, M. (2002). 'Dynamic inconsistency and self-control: a planner-doer interpretation', *Economics Letters,* vol. 77 (3) pp. 419–24.

Benhabib, J. and Bisin, A. (2005). 'Modeling internal commitment mechanisms and self-control: a neuroeconomics approach to consumption-saving decisions', *Games and Economic Behavior,* vol. 52 (2) (August), pp. 460–92.

Bernheim, B. D. and Rangel, A. (forthcoming). 'Behavioral public economics: welfare and policy analysis with fallible decision-makers', in (P. Diamond and H. Vartiainen, ed.), *Economic Institutions and Behavioral Economics,* Princeton: Princeton University Press.

Bhatt, M. and Camerer, C. F. (2005). 'Self-referential thinking and equilibrium as states of mind in games: fmri evidence', *Games and Economic Behavior,* vol. 52 (2) (August), pp. 424–459.

Brocas, I. and Carrillo, J. (2005). 'The brain as a hierarchical organization', University of Southern California.

Brown, A. L., Camerer, C. F. and Chua, Z. E. (2006). 'Learning and visceral temptation in dynamic savings experiments', Caltech.

Bruni, L. and Sugden, R. (2007). 'The road not taken: Two debates on economics and psychology', *Economic Journal*, vol. 117(516), pp. 146–73.

Busino, G. (1964). 'Note bibliographique sur le cours', in (V. Pareto, ed.), *Epstolario,* pp. 1165–72, Rome: Accademia Nazionale dei Lincei.

Camerer, C. F. (2007). *Behavioral Economics,* in (R. Blundell, W. Newey and T. Persson, eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress,* Cambridge: Cambridge Univesity Press.

Camerer, C. F., Ho, T. H. and Chong, J. K. (2004). 'A cognitive hierarchy model of games', *Quarterly Journal of Economics,* vol. 119 (3) (August), pp. 861–98.

Chen, M. K., Lakshminarayanan, V. and Santos, L. (2006). 'How basic are behavioral biases? Evidence from capuchin-monkey trading behavior', *Journal of Political Economy,* vol. 114 pp. 517–37.

Chong, J.-K., Camerer, C. F. and Ho, T. H. (forthcoming). 'A learning-based model of repeated games with incomplete information' *Games and Economic Behavior.*

Chorvat, T. R. and Mccabe, K. (2005). 'Neuroeconomics and rationality', *Chicago-Kent Law Review,* vol. 80 (3) (August), pp. 1235–55.

Colander, D. (2005). 'Neuroeconomics, the hedonimeter, and utility: some historical links', Middlebury College.

Conover, K. and Shizgal, P. (2005). 'Employing labor supply theory to measure the reward value of electrical brain stimulation', *Games and Economic Behavior,* vol. 52 (2), pp. 283–304.

Deaner, R. O., Khera, A. V. and Platt, M. L. (2005). 'Monkeys pay per view: adaptive valuation of social images by rhesus macaques', *Current Biology,* vol. 15 (29 March 2005) pp. 543–8.

Della Vigna, S. and Paserman, M. D. (2005). 'Job search and impatience', *Journal of Labor Economics,* vol. 23 (3) (July), pp. 527–88.

Friedman, M. (1953). *The Methodology of Positive Economics,* Chicago: Chicago University Press.

Fudenberg, D. and Levine, D. (forthcoming). 'A dual self model of impulse control', *American Economic Review.*

Glimcher, P. W., Dorris, M. C. and Bayer, H. M. (2005). 'Physiological utility theory and the neuroeconomics of choice', *Games and Economic Behavior,* vol. 52 (2) (August), pp. 213–56.

Gul, F. and Pesendorfer, W. (2001). 'Temptation and self-control', *Econometrica,* vol. 69 (6) (November), pp. 1403–35.

Gul, F. and Pesendorfer, W. (2005). 'The case for mindless economies', Princeton University, November.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. and Camerer, C. F. (2005). 'Neural systems responding to degrees of uncertainty in human decision-making', *Science,* vol. 310 (5754) (9 December), pp. 1680–3.

Hsu, M., Zhao, C. and Camerer, C. F. (2006). Nonlinear probability weighting in the brain', Caltech.

Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T. and Platt, M. L. (2006). 'Neural signatures of economic preferences for risk and ambiguity', *Neuron,* vol. 49 (5) (2 March 2006), pp. 765–75.

Jevons, W. (1871). *The Theory of Political Economy.* London: Macmillan and Company.

Kagel, J., Battalio, R. C. and Green, L. (1995). *Economic Choice Theory: An Experimental Analysis of Animal Behavior,* Cambridge: Cambridge University Press.

Kahneman, D. (2003). 'A psychological perspective on economics', *American Economic Review,* vol. 93 (2) (May), pp. 162–8.

Kahneman, D. and Tversky, A. (1979). 'Prospect theory – analysis of decision under risk', *Econometrica,* vol. 47 (2) (March), pp. 263–91.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. and Fehr, E. (2005). 'Oxytocin increases trust in humans', *Nature,* vol. 435 (7042) (2 June), pp. 673–6.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. and Fehr, E. (2006). 'Diminishing reciprocal fairness by disrupting the right prefrontal cortex', *Science Express* vol. 314 (5800), pp. 829–32. DOI:10.1126/science.1129156.

Laibson, D. (1997). 'Golden eggs and hyperbolic discounting', *The Quarterly Journal of Economics,* vol. 112 (2) (May), pp. 443–77.

Lee, D., Conroy, M. L., Mcgreevy, B. P. and Barraclough, D. J. (2004). 'Reinforcement learning and decision making in monkeys during a competitive game', *Cognitive Brain Research,* vol. 22 (1) (December), pp. 45–58.

Loewenstein, G. and O'Donoghue, T. (2004). 'Animal spirits: affective and deliberative processes in economic behavior', Carnegie Mellon University,

Mcclure, S. M., Laibson, D. I., Loewenstein, G. and Cohen, J. D. (2004). 'Separate neural systems value immediate and delayed monetary rewards', *Science,* vol. 306 (5695) (15, October), pp. 503–7.

Mullainathan, S. and Thaler, R. (2000). *Behavioral Economics: Entry in International Encyclopedia of the Social and Behavioral Sciences,* Cambridge MA: Massachusetts Institute of Technology.

O'Donoghue, T. and Rabin, M. (1999). 'Doing it now or later', *American Economic Review,* vol. 89 (1) (March), pp. 103–24.

Padoa-Schioppa, C. and Assad, J. A. (2006). 'Neurons in the orbitofrontal cortex encode economic value', *Nature,* vol. 441 (7090) (11 May), pp. 223–6.

Platt, M. L. and Glimcher, P. W. (1999). 'Neural correlates of decision variables in parietal cortex', *Nature,* vol. 400 (6741) pp. 233–8.

Prelec, D. (1998). 'The probability weighting function', *Econometrica,* vol. 66 (3) (May), pp. 497–527.

Raiffa, H. (1961). 'Risk, ambiguity, and the Savage axioms: comment', *Quarterly Journal of Economics,* vol. 75 (4) (Winter), pp. 690–4.

Sanfey, A. G., Loewenstein, G., Cohen, J. D. and Mcclure, S. M. (2006). 'Neuroeconomics: cross-currents in research on decision', *Trends in Cognitive Sciences,* forthcoming.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game', *Science,* vol. 300 (5626) (13 June), pp. 1755–8.

Shefrin, H. M. and Thaler, R. H. (1988). 'The behavioral life-cycle hypothesis', *Economic Inquiry,* vol. 26 (4) (October), pp. 609–43.

Stocker, A. A. and Simoncelli, E. P. (2006). 'Noise characteristics and prior expectations in human visual speed perception', *Nature Neuroscience* vol. 9 (4) (April), pp. 578–85.

Tanaka, T., Camerer, C. F. and Nguyen, Q. (2006). 'Poverty, politics, and preferences: experimental and survey data from Vietnam', California Institute of Technology.

Wang, J. T.-Y., Spezio, M. and Camerer, C. F. (2006). 'Pinocchio's pupil: using eyetracking and pupil dilation to understand truth-telling and deception in biased transmission games', Caltech.

Wout, M. V. T., Kahn, R. S., Sanfey, A. G. and Aleman, A. (2005). 'Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making', *Neurorepor,* vol. 16 (16) (7 November), pp. 1849–52.

Zak, P. J. (2004). 'Neuroeconomics', *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences,* vol. 359 (1451) (29 November 2004), pp. 1737–48.

Zak, P. J., Kurzban, R. and Matzner, W. T. (2005). 'Oxytocin is associated with human trustworthiness', *Hormones and Behavior vol.* 48 (5) (December), pp. 522–7.

TWENTY

The Market as a Creative Process

James M. Buchanan and Viktor J. Vanberg

James M. Buchanan (1919–) is Advisory General Director of the Center for the Study of Public Choice and Harris University Professor at George Mason University. He received his Ph.D. from the University of Chicago, and he is best known for his work in political economy. He was awarded the Nobel Prize for Economics in 1986 in recognition of his work analyzing economic and political decision making.

Viktor J. Vanberg (1943–) is Professor of Economics at George Mason University and coeditor of the journal *Constitutional Political Economy.* He received doctorates from the Technische Universität in Berlin and the University of Mannheim. His research focuses on the economics of institutions and on political economy.

Had Pyrrhus not fallen by a beldam's hand in Argos or Julius Caesar not been knifed to death? They are not to be thought away. Time has branded them and fettered they are lodged in the room of the infinite possibilities they have ousted. But can those have been possible, seeing that they never were? Or, was that only possible which came to pass?

James Joyce[1]

1. Introduction

Contributions in modern theoretical physics and chemistry on the behavior of nonlinear systems, exemplified by Ilya Prigogine's work on the thermodynamics of open systems (Prigogine and Stengers, 1984), attract growing attention in economics (Anderson, Arrow, and Pines, 1988; Arthur, 1990; Baumol and Benhabib, 1989; Mirowski, 1990; Radzicki, 1990). Our purpose

An earlier version of this essay was presented as a paper at a Liberty Fund Conference on "An Inquiry into Liberty and Self-Organizing Systems," April 26–29, 1990, Rio Rico, Arizona. We received helpful comments on previous drafts from Hartmut Kliemt, Karen Vaughn, Jack Wiseman, and an anonymous referee.

Reprinted with the permission of Cambridge University Press from *Economics and Philosophy*, vol. 7 (1991), pp. 167–86.

here is to relate the new orientation in the natural sciences to a particular nonorthodox strand of thought within economics. All that is needed for this purpose is some appreciation of the general thrust of the enterprise, which involves a shift of perspective from the determinism of conventional physics (which presumably inspired the neoclassical research program in economics) to the nonteleological open-endedness, creative, and nondetermined nature of evolutionary processes.

Prigogine and Stengers (1984, p. 177) refer to this shift in perspective as "a reconceptualization of the physical sciences," as a move "from deterministic, reversible processes to stochastic and irreversible ones." The emphasis is shifted from equilibrium to nonequilibrium as a "source of spontaneous self-organization" (Prigogine, 1985, p. 108), to self-organizing processes in open systems far from thermodynamic equilibrium (Prigogine, 1985, p. 108). A characteristic feature of such systems is the presence of nonlinearities that can amplify "small causes" into "large effects." At critical points (referred to as "bifurcations"), very small events can have significant macroeffects, in the sense that they "decide" which particular path – among a number of equally possible paths – the system will take, a fact that introduces a stochastic element and renders self-organizing processes in far-from-equilibrium conditions inherently undetermined.[2] Such processes exhibit a mixture of necessity and chance that, as Prigogine and Stengers note (1984, pp. 169ff.), produces a unique and irreversible "'history' path along which the system evolves."

What is suggested here is a generalized perspective that brings into focus creativity and open-endedness in the evolution of nonequilibrium systems, a perspective that has as its *leitmotiv* "that the future is not given" (Prigogine, 1986, p. 493), but is created in an unfolding evolutionary process.[3] Authors like P. M. Allen (1988, p. 99) and J. S. Wicken (1987, p. 3) speak of a *new evolutionary synthesis*, a "unified view of the world which bridges the gap between the physical and the human sciences" (Allen, 1988, p. 118). In his discussion on the relevance of the "new evolutionary synthesis" for economic theory, Allen stresses the concern with *microscopic diversity* as the critical feature. The "cloudy, confused complexity of the real world" (1988, p. 99) is the essential subject of an evolutionary approach – in contrast to a perspective that looks for types and classes, and that views microscopic diversity and variation as negligible aberrations, to be averaged out through classification and aggregation.[4] Variability and individual diversity at the microscopic level drive evolutionary processes; they are the crucial ingredient to the "creativity" of these processes, of their potential to generate novelty. As Allen (1988, p. 108) puts it: "The fluctuations, mutations and apparently random movements which are naturally present in real complex systems constitute a sort of 'imaginative' and creative force which will explore around whatever exists at present." Allen sees here the critical difference between an evolutionary perspective and one that centers around the notion of predetermined equilibrium states, the difference between the new self-organization paradigm and a "Newtonian paradigm" in which any "representation of 'creative processes' was entirely absent" (*Ibid.*, p. 97).[5]

As noted, our purpose is, first, to identify a body of criticism of orthodox equilibrium theory in economics that seems to correspond closely with the developments noted in the natural sciences, and, second, to elaborate on the implications of this (the *radical subjectivist*) criticism in some detail and, particularly, in its relation to its near neighbor, the entrepreneurial conceptualization of Israel Kirzner.

2. Subjectivism, the Growth of Knowledge, and Indeterminedness

P. M. Allen's article is but one example of the growing number of comments on the apparent relevance of the *new evolutionary synthesis* for a reorientation of economic theory. The reasons that limit the applicability of equilibrium models, even in the traditional realm of physics and chemistry, apply *a fortiori* to the domain of economics. The equilibrium concept is associated with a world view that treats the future as implied in the present. In principle, future states could be predicted based on sufficient knowledge of the present; that is, if it were not for *de facto* limits on our knowledge of an immensely complex reality. By contrast, a core insight of the new paradigm is that nature is creative, that novelty and genuinely unpredictable outcomes are generated as the evolutionary process unfolds over time. The creativity argument has all the more force where concern is with social processes that are driven by human choice and inventiveness.[6]

One criticism of economic orthodoxy that has been advanced from a strict *subjectivist* position (a criticism that has, to our knowledge, been developed independently of the literature discussed above) has, in some respects, a strikingly similar thrust.[7] It should be said at the outset that there is no clearly delineated body of thought that would fall under the rubric of *subjectivism*. The term has been adopted by, and used as a label for, a number of perspectives in economics that agree in their broad criticism of the neoclassical general equilibrium framework, but that are by no means theoretically homogeneous. With this proviso stated, we want to concentrate the discussion here on what is often referred to as "radical subjectivism," a position associated primarily with the name of G. L. S. Shackle (1979) as well as with the work of such other authors as L. M. Lachmann, J. Wiseman, and S. C. Littlechild. In Sec. 3, we shall take a closer look at the modern Austrian version of subjectivism, represented by I. Kirzner's work on entrepreneurship, and we shall discuss the differences that Kirzner sees between his own position and "radical subjectivism."[8]

At the core of Shackle's attack on the "neoclassical citadel" (Lachmann, 1976, p. 54), and central to the radical subjectivist view in general, is the issue of what we can claim to know about the future in our efforts to understand the world of human affairs. The basic objection to neoclassical general equilibrium theory is that it embodies assumptions about the knowability of the future that are entirely unfounded, not only in their most extreme variant, the assumption of perfect knowledge, but also in their softer varieties, such as assumptions about rational expectations or Bayesian adaptive rationality. For radical subjectivism there is simply no way around the fundamental fact that whatever happens in the social realm is dependent on human choices, choices that – if they are *choices* – could be different, and could, if they were different, have different effects.[9] There can, therefore, be no "given" future, independent of the choices that will be made. Instead, there are innumerable potential futures of which only one will emerge as the choice-process unfolds. As Shackle puts it, "the content of time-to-come is not merely unknown but nonexistent, and the notion of foreknowledge of human affairs is vacuous" (1983, p. 33). Or in J. Wiseman's terms: "The essence of the radical subjectivist position is that the future is not simply 'unknown,' but is 'nonexistent' or 'indeterminate' at the point of decision" (1989, p. 230).[10]

The recognition that in human social affairs the future is undetermined but "created" in the process of choice, does not imply that the future is "beyond *conjecture*" (Wiseman, 1990, p. 104), nor does it ignore that individuals have *expectations* about the future on which they base their action. The subjectivist's understanding of the nature and role of such expectations is, however, critically different from their interpretation in a neoclassical framework. To the subjectivist, expectations may be more or less reasonable (in the sense of being more or less defendable in the light of past experience), but they can, ultimately, not be more than conjectures about an undetermined and, therefore, unknowable future. To the neoclassical economist, by contrast, expectations are about a future that is, in principle, *knowable*, even if its knowability may be limited by imperfections of the "expecters." Ignorance of the future is essentially seen as a source of inefficiency, as a problem that can, in principle, be remedied by learning.[11] By contrast, from a subjectivist position, such ignorance is simply "an inescapable characteristic of the human condition" (Wiseman, 1989, p. 225). And "the possibility of learning does not imply that through learning the future will become knowable, but only that experience will change behavior" (*Ibid.*, p. 143).[12]

Arguing on the same theme, Shackle suggests that every person choosing among different courses of action can be seen "to be making history, on however small a scale, in some sense other than mere passive obedience to the play of all-pervasive causes" (1983, p. 28). Every choice can be seen as the beginning of a sequel that "will be partly the work of many people's choices-to-come whose character . . . the chooser of present action cannot know" (*Ibid.*, pp. 28ff.).[13] Our "knowledge" of the future is, from this perspective, not "a deficiency, a falling-short, a failure of search and study" (*Ibid.*, p. 33). Rather, it reflects a fundamental fact of human existence, "the imaginative and originative source and nature of the choosables, and the endless proliferant creation of hypothetical sequels of choosable action" (*Ibid.*, p. 36). It reflects, in other words, "*the plurality of rival possibles*" (*Ibid.*, p. 37).[14]

The emphasis on choice as an *originating* force, the notion of the *creativeness* of the human mind, and the outlook on history as an *open-ended*, evolving process, are intimately interconnected aspects of the same general theme that marks the critical difference between the subjectivist perspective and its neoclassical counterpart. It marks the difference between the *non-teleological* outlook on the human social realm that informs the subjectivist notion of an open-ended, creative-choice process, and the *teleological* thrust that underlies, if only implicitly, the neoclassical notion of an equilibrium solution that is "preordained by patterns of mineral resources, geography, population, consumer tastes and technological possibilities" (Arthur, 1990, p. 99).[15] To Shackle and other radical subjectivists, the whole general equilibrium concept is questionable when applied to a constantly changing social world that has no predeterminable telos, whether in the pompous sense of a Marxian philosophy of history or in the more pedestrian sense of a conceptually definable equilibrium toward which the process of socioeconomic change could be predicted to gravitate. In a world in which creative human choice is a constant source of an "unknowable future," the notion of a "social equilibrium" is, in J. Wiseman's words, a "pseudo-concept" (1989, p. 214), one that can "have only the most tenuous general meaning" (*Ibid.*, p. 265).[16]

Another way of stating the subjectivist objection against the neoclassical equilibrium concept is by saying that the latter does not provide for an adequate account of "real" historical time. It does not take seriously the fact that, as L. M. Lachmann puts it, "*Time* and *Knowledge* belong together" (1977, p. 85), that "time cannot pass without modifying knowledge" (*Ibid.*, p. 93).[17] The common argument that "simplifying assumptions" allow general equilibrium models to ignore the complexities of the "time and knowledge" problem is rejected by Wiseman as unconvincing. The simplifying assumptions about human knowledge are, he argues, "not legitimate simplifications but a gross perversion of the nature of the decision-problem faced by people living in the real world" (1989, p. 140), a defect that cannot be remedied by sophisticated refinements of the models that are based on such assumptions.[18]

The contrast is between two critically different perspectives by which efforts to understand the world can be guided: (1) a *teleological* perspective, and (2) a *nonteleological* perspective. We argue that it is its uncompromising nonteleological character that marks the critical difference between the understanding of the market process suggested by the subjectivist perspective and various standard conceptions of the market that, if only in a very subliminal fashion, have a teleological undertone. And, as an aside, we want to submit that this "residual teleology" constitutes somewhat of a hidden common link between standard economic teaching on the self-organizing nature of markets and the blatant teleology of the socialist planning mentality.

3. Kirzner's Theory of Entrepreneurship

Israel Kirzner's work, with its explicit emphasis on the entrepreneurial role in economic interaction, is of particular interest in the present context because of Kirzner's (1985, pp. 7ff.) claim that his own "alertness" theory of entrepreneurship keeps a balanced middle ground between "two extreme views," the neoclassical equilibrium view on the one side and Shackle's subjectivism on the other, or, in our terms, between a teleological and a nonteleological concept of the market process.[19] As we shall argue, however, in spite of his emphasis on innovative entrepreneurial dynamics and in spite of his verbal recognition of the *creative* and *open-ended* nature of the market process, Kirzner's approach fails to escape the subliminal teleology of the equilibrium framework.[20]

There is, as Littlechild (1979) has pointed out in some detail, a disharmonious mixture in Kirzner's work, between a basic affinity to, and remaining disagreements with, the radical subjectivist position. Kirzner explicitly recognizes the creative dynamics of the market process, and indeed, makes this the central theme of his work. He criticizes the neoclassical position for assigning "*no* role . . . to the creative entrepreneur" (1985, p. 13); he talks of the role of entrepreneurship "in an open-ended, uncertain world" (*Ibid.*, p. 52), a world in which we "find scope for the unpredictable, the creative, the imaginative expression of the human mind" (*Ibid.*, p. 58); and he talks of new products, new qualities of products, new methods of production, and new forms of organization that are endlessly generated in the course of the entrepreneurial process.[21] Yet, such emphasis on creativity, imagination, and novelty is combined with a theoretical perspective that located the essence of entrepreneurship in "the discovery of error" (Kirzner, 1985, p. 50), and the scope for entrepreneurship "in the possibility of discovering error" (*Ibid.*, p. 51), a combination that can hardly be called harmonious.

Discovery of error means, in the context of Kirzner's theory, such things as the discovery of "erroneously low valuation" (*Ibid.*, p. 50) of resources, the "alertness to hitherto unperceived opportunities" (*Ibid.*, p. 52), or the noticing of "situations overlooked until now because of error" (*Ibid.*), phrases that all invite the same questions: If the essence of entrepreneurial discovery is to "provide protection" or "rescue" from "earlier" or "past error" (*Ibid.*, p. 53), what is then the benchmark or *reference-base* against which the failure to do something can be judged to be an "error"? And how does the notion of *creativity* square with such definition of entrepreneurial activity? Are creativity and imagination the same as discovery of errors?

There is, in our view, a fundamental inconsistency in Kirzner's attempt to integrate the innovativeness of entrepreneurial activity into an equilibrium framework – by modeling it as *discovery* of "erroneously overlooked opportunities."[22] The critical step in Kirzner's argument, the step that is intended to establish a "middle ground" between a teleological and a non-teleological understanding of the market process, is his extension of the notion of a divergence between "different parts of the market" (1985, p. 62) from a *cross-sectional* to an *intertemporal* interpretation.[23] According to the cross-sectional interpretation, the entrepreneur acts essentially as *arbitrageur:* By taking advantage of hitherto unnoticed divergences between different parts in a present market, he helps to bring about greater consistency (Kirzner, 1985, pp. 61ff.). According to the intertemporal interpretation, the entrepreneur takes advantage of yet unnoticed divergences between *today's* market and *tomorrow's* market, thus helping "to coordinate markets also across time" (*Ibid.*, p. 62).[24]

Whatever may be said about the knowability of divergences in the cross-sectional interpretation, it should be obvious that the notion of *intertemporal* divergences between markets at different points in time is inherently problematic. If, as we must assume, divergences between today's and tomorrow's markets are typically associated with differences between today's and tomorrow's *knowledge*, what does it mean to say that entrepreneurial alertness corrects the "failure to realize" divergences between *present* and *future* markets? What sense does it make to describe today's failure to possess tomorrow's knowledge as *error*?[25] If, to use Lachmann's phrase, "*Time* and *Knowledge* belong together," a comparison between present and future markets cannot possibly be made in a sense that would make such terminology meaningful. The kind of comparison that can be made, at least conceptually, across contemporaneous markets cannot be made along the "intertemporal dimension" (Kirzner, 1985, p. 62). Time is not simply another "dimension," comparable to the spatial. Different parts of a present market exist, they are *present*, and differences in their characteristics can be discovered. Future parts of a market simply do not exist; they are, by definition, not present. There are, at any point in time, many *potential* futures imaginable, based on more or less informed reflections. Yet, which future will come into existence will depend on choices that are yet to be made. Of course, human beings aim to be "prepared for the future," and they act on their expectations of what lies ahead. The subjectivist argument on the unknowability of the future is certainly not meant as a recommendation to merchants not to anticipate the coming of winter in their storekeeping. Yet, if, and to the extent that, human choices and their complex interactions shape the emerging future, the latter can be a matter of speculation, but not of foreknowledge.

The supposition that the future is foreknowable clearly seems implied when, in talking about the problem of intertemporal entrepreneurial alertness, Kirzner speaks of pictures of the future that may or may not "correspond to the truth as it will be realized" (1985, p. 55), of man's efforts to overcome uncertainty "by more accurate prescience" (*Ibid.*, p. 58), of "past failure to pierce correctly the fog of uncertainty" (*Ibid.*, p. 53), and so forth. It is far from obvious how such insinuation of a preknowable future can be consistent with a genuine appreciation of the creativity of the human mind. Indeed, when arriving at this issue, Kirzner simply retreats to the *ex cathedra* claim that his approach does encompass the two notions, without actually showing *how* this can be done. He emphasizes that intertemporal entrepreneurial alertness "does not consist merely in 'seeing' the unfolding of the tapestry of the future in the sense of seeing a preordained flow of events" (1985, p. 56). Indeed, he insists that such alertness must "embrace the awareness of the ways in which the human agent can . . . in fact *create* the future" (*Ibid.*). Yet, as if the compatibility of the two arguments were obvious, he also insists that the function of market entrepreneurship in the multiperiod context is nonetheless still that of "discovery of errors" in the sense explained above (*Ibid.*).[26] And he leaves undiscussed the issue of what one entrepreneur's creativity means for the truthfulness of another entrepreneur's picture of the future.[27]

If, as Kirzner's construction seems to suggest, today's failure to possess tomorrow's knowledge qualifies as *error* from which entrepreneurial alertness is to provide rescue, one could conclude that the ultimate benchmark or reference base for such judgment is an imagined world in which everything that humans may ever imagine, think, or know will be revealed.[28] Judged against such a benchmark, every act, however imaginative and creative, can be seen as a discovery of something that was already waiting to be found. And failure to discover may be discussed in terms of error and overlooked opportunities. It seems questionable, however, whether the mental construct of such an imagined world is a helpful analytical guide when applied to the study of socioeconomic change.

What might be misleadingly suggestive here is the analogy to the scientific discovery process. To the extent that science is concerned with an objective reality "out there," our conjectural knowledge of this reality can be expected to grow over time, through a process of discovery. Although we cannot know at present what we will know in the future, any future increase in knowledge can, in some sense, be viewed as a finding of something that could, in principle, be currently discovered. There is something knowable out there, to be discovered sooner or later. Any such account of the discovery process in science is itself seriously challenged by the new conceptions advanced by Prigogine and others, because of its neglect of real time. But, even if, for the purpose of our discussion here, we should leave this issue aside, the analogous challenge advanced by the radical subjectivists to neoclassical equilibrium economics applies with full force to the concept of the market as a discovery process. Entrepreneurial activity, in particular, is not to be modelled as discovery of that which is "out there." Such activity, by contrast, *creates* a reality that will be different subsequent on differing choices. Hence, the reality of the future must be shaped by choices yet to be made, and this reality has no existence independent of these choices. With regard to a "yet to be created" reality, it is surely confusing to consider its emergence in terms of the discovery of "overlooked opportunities."[29]

4. Conceptions and Misconceptions of the Market

The essential characteristic of the radical subjectivist position that marks its critical departure from a neoclassical framework is, at the same time, the feature that it shares with the new evolutionary synthesis discussed at the beginning of this article: Its conception of "a world in which time plays a vital role" (Littlechild, 1979, p. 38), of history as an open-ended evolving process, and of a future that is not predetermined, merely waiting to be revealed, but that is "continuously *originated* by the pattern and sequence of human choice" (*Ibid.*). Such a conception has clear implications for the theory of the market that set it apart from various theoretical constructs that have been used to explain or to illustrate the adaptive nature of the market process. If the emphasis on the creativity of human choice is taken seriously, it is not only the standard neoclassical equilibrium notion that seems questionable, but also less orthodox conceptions of the market process, including Kirzner's more subliminally teleological perspective on markets and entrepreneurship. By stating this we certainly do not want to suggest that "radical subjectivism" exists as a well-specified theoretical paradigm ready for adoption – it clearly is not. What we want to suggest, however, is that the creativity of human choice poses a problem that any effective socioeconomic theory cannot evade.

The critical shift in perspective may be further illustrated by reference to three separate understandings of the spontaneous order of the market that have been advanced by scholars who have been generally supportive of market organization of the economy, no one of whom would ever have referred to the market as an "analogue computer" for the "computation of equilibrium prices."

1. One of us (Buchanan) learned basic price theory at the University of Chicago in the 1940s, when all students, undergraduate and graduate, were required to master the Syllabus written by Henry Simons.[30] This Syllabus contained three well-known rent problems that were designed to provide an understanding of how a competitive economy allocates scarce resources among uses. And, as a test of the efficacy of competitive adjustment, one task given to the students was that of comparing the total product of the economy in competitive equilibrium with that which might be achieved under allocation by a benevolent and omniscient planner.

2. In a deservedly famous article, "The Logic of Liberty," Michael Polanyi introduced the metaphor of a sack of potatoes that need only to be shaken to insure minimization of volume to demonstrate how localized, decentralized adjustment, akin to that which is characteristic of market organization, works better than centralized adjustment.[31]

3. In a monograph-length essay devoted to an explication of the spontaneous order of the market, Norman Barry (1982) stated that the results of a market "appear to be a product of some omniscient, designing mind."[32]

In each of these illustrative examples, there is revealed, at least by inference, an understanding of the spontaneous ordering properties of a market process that is sharply different from the understanding held by the radical subjectivists. In each example, the efficacy of market adjustment is measured *teleologically* in terms of the relative achievement of some predefined goal or objective. In Simons' problems, the objective is, simply, economic product, which is wheat in his one-good economy. In Polanyi's case, the objective is explicitly stated to be minimization of volume. In Barry's essay, the argument is more sophisticated, but any conceptualization of an omniscient, designing mind must imply some well-defined objective that exists independently from the separate participants' own *creative* choices.

If the efficacy of market organization, is, as insinuated in the above examples, evaluated teleologically, in terms of its capacity to approach an independently (that is, independent of the choice of process itself) determinable state, then there remains only an ambiguous discourse over comparative performance as between such an organization and centralized economic planning. Even if Simons, Polanyi, and Barry, along with others, may have succeeded in demonstrating that decentralized arrangements are superior in achieving some objectively identifiable goal, their conceptualization of the market process forces them into a line of comparative defense that a radical subjectivist understanding of the market would have rendered unnecessary from the outset. If the market is genuinely perceived as an open-ended, nondetermined evolutionary process in which the essential driving force is human choice, any insinuation, however subtle, of a "telos" toward which the process can be predicted to move must be inherently misleading. There is, in our view, no systematically sustainable middle ground between a teleological and a nonteleological perspective. And all conceptualizations of the market process that suppose, whether explicitly or implicitly, a "something" toward which the process is moving are, by this very fact, *teleological*, whether the "something" is specified as an equilibrium or otherwise. This applies to the notion of a mechanical equilibrium as implied in the standard textbook models of intersecting demand and supply curves, as well as to the thermodynamic equilibrium concept that is implied where the market process is interpreted in terms of exhaustion of potential gains from trade. And it also applies to images of the market that are intended to capture the constant change in the equilibrium-telos, such as K. Boulding's image of the "dog chasing a cat" (Littlechild, 1986, p. 32).

It should be noted that to question the appropriateness of teleological conceptions of the market is not the same as denying the apparent fact that the human participants in the "catallaxy," the game of the market, reasonably *adapt* to the circumstances that they confront and to changes that they expect to occur. The predictive potential of microeconomic theory lies in the uniformity of such adaptive response among persons. But such adaptive behavior does not imply that the overall process is moving toward some determined goal, whether conceived as a predetermined equilibrium or as a "moving cat." The game described by the market may be misunderstood if interpreted in a teleological mind-set. The market economy, *as an aggregation*, neither maximizes nor minimizes anything. It simply allows participants to pursue that which they value, subject to the preferences and endowments of others, and within the constraints of general "rules of the game" that allow, and provide incentives for, individuals to try out new ways of doing things. There simply is no "external," independently defined objective against which the results of market processes can be evaluated.

We may illustrate the nonteleological perspective on market interaction by dropping the familiar presupposition that potential traders initially possess quantities of well-defined marketable goods. Assume that no goods exist, and that persons are described by certain talents, capacities, and skills that enable them to produce consumable goods from nature. Assume that the rules of the game allow persons to claim enforceable rights to the shares in natural endowments and to their own capacities and skills. In this model, trade will take place when persons recognize that their well-being can be enhanced by producing *and* exchanging rather than producing for their own consumption only. But the chain of choices is extended, and, also, there is an added requirement that any participant exercise *imagination* in choosing to specialize in production with the ultimate purpose of achieving an increase in well-being through exchange.

Think of the choice calculus of a person in this setting. What can I produce that will prove of exchange value to others? Response to this question allows the participant not only to select among a preexisting set of goods, but, also and importantly, to *create* new goods that are expected to be of potential exchangeable value. Once the creative-inventive-imaginative element in choice is introduced into the game here, then any idealized omniscience on the part of a planner who might attempt to duplicate the market result would become patently absurd. Individuals would use their own imagination, their own assessment of the potential evaluations of others, in producing goods wholly divorced from their own consumption, goods that are anticipated to yield values when put on the market, values that, as income to the producers, can be used to purchase goods from others in the nexus. This seeking to satisfy others through producing marketable value as an indirect means of producing value for themselves – this characteristic behavioral element in a market order was central to Adam Smith's insight. And it is this feature that allows us to compare the performance of market organization with alternative social arrangements, even in the absence of an independently existing scalar. Markets tend to satisfy the preferences of persons, regardless of what their preferences might be, and even when we acknowledge that preferences emerge only within the process of choice itself.

The market conceived as a “game without goods” also suggests the tenuousness of the whole notion of equilibrium, defined as the exhaustion of gains from trade, which looms so important in the alternative teleological perspective. In the production and exchange of preexisting and well-defined goods, it is relatively easy to think of the game as having a definitive and final outcome once the goods have been so allocated that no participant seeks out further trades. Goods are, by definition, then allocated to their highest valued uses. But the usefulness of this equilibrium notion becomes less clear when we assume that there is no definite set of goods to be allocated. Conceptually, it remains possible to “freeze” the imaginative elements in individual choice at some point and allow the production-exchange process to work itself out to an equilibrium, where no further gains from trade, *and from imagination of new trading prospects*, are possible. The artificiality of such an equilibrium construction is apparent, however, since there seems nothing in the mind that is even remotely analogous to the cessation of exchange. There is no determinate limit to the potential of market value to be created as the process of human interaction proceeds.

What has made, and continues to make, the equilibrium concept attractive even to economists who, like Kirzner, are explicitly critical of the neoclassical orthodoxy is, it seems, its perceived capacity to readily capture the coordinative properties of markets, and the suspicion that the radical subjectivist critique may leave one incapable of systematically accounting for the orderliness of markets. Even if such suspicion may have been invited by some of the radical subjectivists, the emerging *new evolutionary synthesis* suggests a theoretical perspective that allows the subjectivist emphasis on the creativity of human choice, with all its implications, to be taken seriously, while, at the same time, it offers nonteleological explanations for the adaptiveness and coordinative properties that markets exhibit.

5. Conclusion

We have suggested that a perceptual vision of the market as a *creative process* offers more insight and understanding than the alternative visions that elicit interpretations of the market as a *discovery process*,[33] or, more familiarly, as an *allocative process*. In either of the latter alternatives, there is a telos imposed by the scientist’s own perception, a telos that is nonexistent in the first stance. And removal of the teleological inference from the way of looking at economic interaction carries with it significant implications for any diagnosis of failure or success, diagnosis that is necessarily preliminary to any normative usage of scientific analysis.

We may illustrate the differing implications in application to the observed failure of the centrally planned economies of Eastern Europe and elsewhere. The neoclassical economist, trapped in the allocationist perception, tends to locate the source of failure in the distorted incentive structure that causes persons to be confronted with choice alternatives that do not reflect authentically derived evaluations. Resources do not flow to their most highly valued uses because persons who make decisions about resource use do not find it privately in their own interest to shift allocation in such fashion as to accomplish this conceptually definable, and desirable, result.

Some of the modern Austrian economists, and notably Kirzner, add an important element to the neoclassical critique. They suggest that, even if the incentive problems could, somehow, be ignored or assumed corrected, there would still remain the epistemological or knowledge problem. Only a decentralized market structure of economic interaction can exploit fully the knowledge of localized circumstances required to allow a definition of the ultimate valuation that is placed on resource use. Only the market can allow persons the effective liberty to discover the particular localized eccentricities that give form to value. This extension of the neoclassical emphasis on incentive structures is important and relevant to any overall assessment of the central planning model for an economy.

We suggest, however, that the critique, even as extended, falls short of capturing an essential element in any comparative assessment of the market and the planning alternatives. The teleological feature remains to be exorcised. In the neoclassical setting, even as extended by Kirzner, an *omniscient* and *benevolent* monolithic planner could secure the ideally defined result. Omniscience would, of course, insure access to any and all knowledge; benevolence could be such as to match the objective function precisely with whatever it is that individuals desire. But even the planner so idealized cannot create that which is not there and will not be there save through the exercise of the creative choices of individuals, who themselves have no idea in advance concerning the ideas that their own imaginations will yield.

The fundamental misunderstandings of the theory of the market economy that provided the analytical-intellectual foundations for socialism as a principle for socioeconomic organization are exposed by any one of the three interpretations contrasted here. The market as an allocative process, responding to the structure of incentives that confront choice makers; the market as a discovery process, utilizing localized information; or the market as a creative process that exploits man's imaginative potential – socialism cannot, organizationally, be made equivalent to any one of these idealized perceptions. But, the "fatal conceit" that was socialism, to use Hayek's descriptive term here, would have surely faced more difficulty in achieving dominance as an idea if the creative spontaneity of the market process had been more fully appreciated.

Notes

1. Joyce, 1960. p. 30.
2. Prigogine and Stengers: "Whenever we reach a bifurcation point, deterministic description breaks down. The type of fluctuation present in the system will lead to the choice of the branch it will follow. Crossing a bifurcation point is a stochastic process, such as the tossing of a coin" (1984, p. 177).
3. Prigogine: "[W]e come to a world which is open, in which the past is present and cumulative, in which the present is there but the future is not. . . . The future does not exist yet, the future is in construction, a construction which is going on in all existing activities" (1985, p. 117).
4. The critical importance of individual diversity and variation from an evolutionary perspective is similarly stressed by biologist E. Mayr, who uses in this context the term "population thinking": "Population thinkers stress the uniqueness of everything in the organic world. What is important for them is the individual, not the type. . . . There is no 'typical' individual, and mean values are abstractions. . . . The differences between biological individuals are real, while the mean values which we may calculate in the comparison of groups of individuals (species, for example) are man-made inferences" (Mayr, 1982, pp. 46ff.). Mayr contrasts "population thinking" with "essentialist thinking": "Adoption of population thinking is intimately tied up with a rejection of essentialist thinking. Variation is irrelevant and therefore uninteresting to the essentialist. Varying characters are 'mere accidents,' in the language of essentialism" (*Ibid.*, p. 487).
5. As P. Allen points out, one has to realize "that there is a critical difference between asking whether a system *obeys* the laws of physics, . . . or whether its behavior can be predicted from a knowledge of those laws" (1985, pp. 268ff.). For nonlinear systems, Allen argues, the first can be the case without the second being possible, due to the mixture of deterministic and stochastic aspects of nonlinear systems (*Ibid.*, p. 270). Allen's argument parallels K. R. Popper's remark in *The Open Universe:* "[C]ausality has to be distinguished from determinism, and our world of uniqueness is – unlike Kant's noumenal world – in space and, even more important, in time; for I find it crucially important to distinguish between the determined *past* and the open *future*" (1982, p. 48). In reference to Prigogine's work, Popper argues in the same treatise: "We must not . . . blind us to the fact that the universe that harbours life is creative in the best sense: creative in the sense in which the great poets, the great artists, the great musicians have been creative, as well as the great mathematicians, the great scientists, and the great inventors" (*Ibid.*, p. 174).
6. Prigogine: "Clearly, a social system is by definition a nonlinear one, as interactions between the members of the society may have a catalytic effect. At each

moment fluctuations are generated, which may be damped or amplified by society. An excellent example of a huge amplification . . . is the acquisition of knowledge. . . . Instead of seeing human systems in terms of 'equilibrium' or as a 'mechanism,' we see a creative world of imperfect information and shifting values, in which different futures can be envisaged" (1986, p. 503).

7. This similarity has been explicitly noted by Fehl (1986); see also Witt (1985).
8. There are other versions of "economic subjectivism" that can be distinguished from both its "radical" and Austrian variety, in particular, the "opportunity costs approach" that has been systematically stated by one of the present authors (Buchanan, 1969, 1987). This version, as well as others that could be identified, will, however, not be discussed as such in the present article.
9. Allen: "The response to this question of 'choice,' which makes modelling and predicting difficult, can be of two kinds. Either we can suppose that choice is an illusion and that the mechanical analogy is in fact legitimate, or we must find some new scientific paradigm in which 'choice' really exists" (1985, p. 269).
10. Littlechild stresses that same point when he summarizes the "radical subjectivst" view as implying that the "as-yet-undetermined actions of other agents" make for "the essential open-endedness of creativity" (1986, p. 31) in human affairs, that "the future is not so much unknown as it is nonexistent or indetermined at the time of decision" (*Ibid.*, p. 29).
11. Wiseman: "Mainstream economics deals with unknowability by assuming it away. In the simple model, this is done by assuming perfect knowledge of the future. . . . The more sophisticated models assume knowledge of the possible number of future states of the world. . . . They assume that *someone* has a knowledge of the future that no one can possibly have" (1990, p. 103). See also Wiseman (1989, p. 159).
12. Wiseman: "*The future* has not yet happened. About it, men can have only *opinions*, related to past experience (learning). Since men can (must) choose how to act, their chosen acts, together with the evolution of the physical world, are continuously creating the emerging future. If this is so (as it must be), then the future cannot be known 'now' (that is, in the continuous present)" (1989, p. 268).
13. As a summary of Shackle's position, Littlechild states, "Choice . . . represents an origin, a beginning. . . . [I]t does have a sequel. It makes a difference to what comes after. This sequel cannot be foreknown, because subsequent events will depend partly upon other such choices yet to be made" (1979, p. 33).
14. Shackle: "[I]f we had *all the data there are or could be* about the *present*, we might still not be able to infer what the sequel of any action now chosen would be. . . . If history, past and to come, is all one book already written at the beginning of time, what is choice? . . . But if choice is fertile, effective, truly *inceptive*, then there can be no foreknowledge. History-to-come, in that case, is not only unknown but *not yet existent*" (1981, p. 60).
15. We use the term "teleological" here in a more general sense than that of an explanation in terms of intended ends or purposeful design. We classify as "teleological" all theoretical perspectives that explain processes in terms of some predeterminable end point toward which they are supposed to move, rather than in terms of explicitly specified forces and principles that actually "drive" them. It is in this sense that we classify as "teleological" an equilibrium theory that

describes economic processes in terms of "where they are going," namely, their end-point equilibria, but does not provide an explicit explanatory account of the dynamics of these processes themselves.

16. Littlechild: "[F]or G. L. S. Shackle, the relevance of the whole concept (of general equilibrium) is in question. Every act of choice embodies the chooser's creative imagination of the future. The market therefore follows a 'kaleidic' process, with moments of order interspersed with disintegration into a new pattern. The economy is changing and developing, but in no sense does it have a single goal" (1983, pp. 48ff.).
17. Lachmann: "The impossibility of prediction in economics follows from the fact that economic change is linked to change in knowledge, and future knowledge cannot be gained before its time. Knowledge is generated by spontaneous acts of the mind" (1977, p. 90).
18. Wiseman: "But if what is assumed away is the essence of the problem, then greater complexity will generate not greater insights but more sophisticated confusion" (1989, p. 227).
19. Kirzner: "I claim, indeed, that the 'alertness' view of entrepreneurship enables us to have the best of both worlds: we *can* incorporate entrepreneurship into the analysis without surrendering the heart of microeconomic theory" (1985, p. 11). Stated differently, Kirzner claims to avoid the neoclassical orthodoxy's failure to account for "the creative entrepreneur" (*Ibid.*, p. 13), without falling "into the seductive trap offered by the opposite extreme" (*Ibid.*), that is, by the radical subjectivist position.
20. G. P. O'Driscoll's and M. J. Rizzo's exposition of a modern Austrian-subjectivist economics is, in a similar way, characterized by a tension between the acceptance of basic tenets of radical subjectivism and the attempt to maintain "an appropriately revised idea of equilibrium" (1985, p. 79).
21. Kirzner: "In the course of this entrepreneurial process, new products may be introduced, new qualities of existing products may be developed, new methods of production may be ventured, new forms of industrial organization, financing, marketing, or tackling risk may be developed. All the ceaseless churning and agitation of the market is to be understood as the consequence of the never-ending discovery process of which the market consists" (1985, pp. 30ff.).
22. Kirzner: "I postulate a continuous discovery process – an entrepreneurial discovery process – that in the absence of external changes in underlying conditions, fuels a tendency toward equilibrium" (1985, p. 12).
23. Kirzner: "What market entrepreneurship accomplishes is a tendency for transactions in different parts of the market (including the market at different dates) to become coordinated" (1985, p. 64).
24. Kirzner's crucial argument, in this context, is worth quoting at some length: "When we introduce the passage of time, the dimensions along which mutual ignorance may develop are multiplied. Market participants in one part of today's market may not only be imperfectly aware of the transactions available in another part of the market; they also may be imperfectly aware of the transactions that will be available in next year's market. Absence of consistency between parts of today's market is seen as a special case of a more general notion of inconsistency that includes also inconsistency between today's transactions and those to be

transacted next year. . . . It is still the case, as noted, that the entrepreneurial function is that of bringing about a tendency for transactions in different parts of the market (conceived broadly now as including transactions entered into at different times) to be made in greater mutual consistency. But whereas in the case of entrepreneurship in the single-period market (that is, the case of the entrepreneur as arbitrageur) entrepreneurial alertness meant alertness to present facts, in the case of multiperiod entrepreneurship alertness must mean alertness to the future" (1985, pp. 62ff.).

25. A well-known classical statement of the argument that we simply cannot anticipate future knowledge and, therefore, cannot predict future human choices that will be affected by such future knowledge, can be found in K. R. Popper's Preface to his *The Poverty of Historicism* (1957).
26. The same kind of tension between Kirzner's chosen theoretical framework and his attempt to incorporate the notion of entrepreneurial inventiveness in the creation of new products and new ways of doing things is also visible in his more recent discussion on the subject (Kirzner, 1989, pp. 84ff.). In her review of this book, K. Vaughn comments on Kirzner's attempts to account for the creative aspects of entrepreneurship while retaining his earlier language: "It has become obvious to this reviewer that the old language no longer fits his new theoretical insights" (1990, p. 185).
27. Kirzner indirectly refers to this issue without, however, discussing it: "In particular the futurity that entrepreneurship must confront introduces the possibility that the entrepreneur may, by his own creative actions, in fact *construct* the future as *he* wishes it to be. In the single-period case alertness can at best discover hitherto overlooked current facts. In the multiperiod case entrepreneurial alertness must include the entrepreneur's perception of the way in which creative and imaginative action may vitally shape the kind of transactions that will be entered into in future market periods" (1985, pp. 63ff.).
28. And, by implication, one could argue that the "equilibrium" toward which intertemporal coordination – as it is promoted by entrepreneurial discovery of error – tends to gravitate can only be some final state of universal enlightenment, at the end of all times. Support for such, admittedly exaggerated, interpretation may be seen in statements such as this: "My view, therefore, sees initial market ignorance indeed as an inescapable feature of the human condition in a world of change, but also as subject to continual erosion. . . . (Entrepreneurs) discover where existing decisions were in fact mistaken. Here lies the source for any equilibrating tendencies that markets display" (Kirzner, 1985, p. 13).
29. The discussion here, and elsewhere in this article, is related, at least indirectly, to a criticism of Michael Polanyi advanced by one of us in two related articles (Buchanan, 1977, 1985). Polanyi conceptualized the scientific process as exploration or discovery, and he argued persuasively that decentralized organization of the scientific enterprise would insure more rapid advance in "solving" the "jigsaw puzzle." From this conceptualization of the scientific process, Polanyi supported, by analogy, the spontaneous ordering properties of decentralized market processes.

    Buchanan's criticism suggested that, even if the discovery-exploration metaphor remains applicable to the enterprise of the physical sciences, such a metaphor is misleading when applied and extended to economic or political interaction among freely choosing individuals.

30. The Simons' Syllabus was circulated only in mimeographed form. Gordon Tullock, himself a student of Simons in the 1940s, edited and published a somewhat incomplete version in 1983 (Tullock, 1983).
31. This article was the title essay in the volume *The Logic of Liberty* (Polanyi, 1951).
32. For a commentary on Barry's essay, see Buchanan (1982).
33. Although the thrust of his work clearly supports the vision of the market as a creative process, Hayek's (1978) illuminating discussion on "Competition as a Discovery Procedure" is not entirely free of the ambiguities that the concept of *discovery* tends to invoke when applied to the market process. Potentially misleading are, in this regard, his comparison between the discovery processes in science and in the market (*Ibid.*, p. 181) and some of his comments on the problem of measuring market performance (*Ibid.*, pp. 185ff.).

References

Allen, Peter M. 1985. "Towards a New Science of Complex Systems." In *The Science and Praxis of Complexity*, by S. Aida et al. Tokyo: The United Nations University, pp. 268–97.

———. 1988. "Evolution, Innovation and Economics." In *Technical Change and Economic Theory*, ed. G. Dosi, C. Freeman, R. Nelson, G. Silverberg, and L. Soete. London: Pinter Publishers Ltd., pp. 95–119.

Anderson, Philip W., Kenneth J. Arrow, and David Pines (editors). 1988. *The Economy as an Evolving Complex System.* New York: Addison-Wesley.

Arthur, W. Brian. 1990. "Positive Feedbacks in the Economy." *Scientific American* 262: 92–9.

Barry, Norman. 1982. "The Tradition of Spontaneous Order." *The Literature of Liberty* 5:7–58.

Baumol, William, and Stephen Benhabib. 1989. "Chaos: Significance, Mechanism, and Economic Applications." *Journal of Economic Issues* 3:77–106.

Buchanan, James M. 1969. *Cost and Choice: An Inquiry in Economic Theory.* Chicago: Markham Publishing Company.

———. 1977. "Politics and Science." In *Freedom in Constitutional Contract*, by J. M. Buchanan. College Station: Texas A&M University Press, pp. 64–77.

———. 1982. "Order Defined in the Process of Its Emergence." *The Literature of Liberty* 5:5.

———. 1985. "The Potential for Tyranny in Politics as Science." In *Liberty, Market and State*, by J. M. Buchanan. New York: New York University Press, pp. 40–54.

———. 1987. "L. S. E. Cost Theory in Retrospect." In *Economics: Between Predictive Science and Moral Philosophy*, by J. M. Buchanan. College Station: Texas A&M University Press, pp. 141–51.

Fehl, Ulrich. 1986. "Spontaneous Order and the Subjectivity of Expectations: A Contribution to the Lachman-O'Driscoll Problem." In *Subjectivism, Intelligibility, and Economic Understanding*, ed. I. M. Kirzner. New York: New York University Press, pp. 72–86.

Hayek, Friedrich A. 1978. "Competition as a Discovery Procedure." In *New Studies in Philosophy, Politics, Economics, and the History of Ideas*, by F. A. Hayek. Chicago: The University of Chicago Press, pp. 179–90.

Joyce, James. 1960. *Ulysses*. London: Bodley Head.

Kirzner, Israel M. 1985. *Discovery and the Capitalist Process*. Chicago: The University of Chicago Press.

———. 1989. *Discovery, Capitalism, Distributive Justice*. New York: Basil Blackwell.

Lachmann, Ludwig, M. 1976. "From Mises to Shackle: An Essay on Austrian Economics and the Kaleidic Society." *Journal of Economic Literature* 14:54–62.

———. 1977. "Professor Shackle on the Economic Significance of Time." In *Capital, Expectations, and the Market Process*, by L. M. Lachmann. Kansas City, MO: Sheed Andrews and McMeel, pp. 81–93.

Littlechild, Stephen C. 1979. "Comment: Radical Subjectivism or Radical Subversion." In *Time, Uncertainty and Disequilibrium: Exploration of Austrian Themes*, ed. M. Rizzo. Lexington, Mass.: Lexington Books, pp. 32–49.

———. 1983. "Subjectivism and Method in Economics." In *Beyond Positive Economics*, ed. J. Wiseman. London: Macmillan, pp. 38–49.

———. 1986. "Three Types of Market Process." In *Economics as a Process – Essays in the New Institutional Economics*, ed. Richard N. Langlois. Cambridge: Cambridge University Press, pp. 27–39.

Mayr, Ernst. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, Mass.: Harvard University Press.

Mirowski, Philip. 1990. "From Mandelbrot to Chaos in Economic Theory." *Southern Economic Journal* 57:289–307.

O'Driscoll, Gerald P., and Mario J. Rizzo. 1985. *The Economics of Time and Ignorance*. New York: Basil Blackwell.

Polanyi, Michael. 1951. *The Logic of Liberty*. Chicago: The University of Chicago Press.

Popper, Karl R. 1957. *The Poverty of Historicism*. Boston: The Beacon Press.

———. 1982. *The Open Universe: An Argument for Indeterminism*. Totowa, N.J.: Rowan and Littlefield.

Prigogine, Ilya. 1985. "New Perspectives on Complexity." In *The Science and Praxis of Complexity*, by S. Aida et al., Tokyo: The United Nations University, pp. 107–18.

———. 1986. "Science, Civilization and Democracy." *Futures* 18:493–507.

Prigogine, Ilya, and Isabelle Stengers. 1984. *Order out of Chaos: Man's New Dialogue with Nature*. Toronto: Bantam Books.

Radzicki, Michael J. 1990. "Institutional Dynamics, Deterministic Chaos, and Self-organizing Systems." *Journal of Economic Issues* 24:57–102.

Shackle, G. L. S. 1979. *Imagination and the Nature of Choice*. Edinburgh: Edinburgh University Press.

———. 1981. "Comments." In *Subjectivist Economics: The New Austrian School*, by A. H. Shand. Oxford: The Pica Press, pp. 59–67.

———. 1983. "The Bounds of Unknowledge." In *Beyond Positive Economics*, ed. J. Wiseman. London: Macmillan, pp. 28–37.

Tullock, Gordon (editor). 1983. *The Simons' Syllabus*, by Henry Calvert Simons. Blacksburg: Virginia Polytechnic Institute and State University.

Vaughn, Karen I. 1990. "Profits, Alertness and Imagination" (review of I. M. Kirzner's *Discovery, Capitalism, and Distributive Justice). Journal des Economistes et des Etudes Humaines* 1:183–8.

Wicken, Jeffrey S. 1987. *Evolution, Thermodynamics, and Information: Extending the Darwinian Paradigm.* Oxford: Oxford University Press.

Wiseman, Jack. 1989. *Cost, Choice, and Political Economy*. Aldershot: Edward Elgar.

———. 1990. "Principles of Political Economy: An Outline Proposal, Illustrated by Application to Fiscal Federalism." *Constitutional Political Economy* 1:101–27.

Witt, Ulrich. 1985. "Coordination of Individual Economic Activities as an Evolving Process of Self-Organization." *Economie Appliquée* 37:569–95.

TWENTY-ONE

What Is the Essence of Institutional Economics?

Geoffrey M. Hodgson

Geoffrey M. Hodgson (1946–) is a Research Professor in Business Studies at the University of Hertfordshire (UK), the Editor-in-Chief of the *Journal of Institutional Economics* and was 2006 President of the Association for Evolutionary Economics. He is the author of more than a dozen books and nearly two hundred scholarly articles. His research has focused on institutions, and he also has had a long-standing interest in the history and methodology of institutional and evolutionary economics.

The term "institutional economics" was announced by Walton Hamilton at a meeting of the American Economic Association in 1918 [Hamilton 1919]. Institutionalism dominated American economics, at least until the 1940s. Listing a number of perceived attributes of this school, Walton Hamilton [1919, 309–11] claimed that institutional economics alone could unify economic science by showing how parts of the economic system related to the whole. Institutional economics was not defined in terms of any normative stance. Hamilton [1919, 313] declared: "It is not the place of economics to pass judgments upon practical proposals." However, its appeal as a theory was that allegedly it could be used as a basis for policy. According to Hamilton [1919, 314–18], institutional economists recognized that:

The proper subject-matter of economic theory is institutions. . . . Economic theory is concerned with matters of process. . . . Economic theory must be based upon an acceptable theory of human behavior . . .

This was expanded by the following observations:

neo-classical economics . . . neglected the influence exercised over conduct by the scheme of institutions . . . Where it fails, institutionalism must strive for success . . . it must discern in the variety of institutional situations impinging upon individuals the chief source of differences in the content of their behavior [1919, 318].

Journal of Economic Issues, vol. 34 (June 2000): 317–29. Reprinted by special permission of the copyright holder, the Association for Evolutionary Economics.

Hamilton's description of institutionalism requires refinement, but in its essentials it has endured the test of time. It can be rephrased and expanded in terms of the following five propositions:

1. Although institutional economists are keen to give their theories practical relevance, institutionalism itself is not defined in terms of any policy proposals.
2. Institutionalism makes extensive use of ideas and data from other disciplines such as psychology, sociology and anthropology in order to develop a richer analysis of institutions and of human behavior.
3. Institutions are the key elements of any economy, and thus a major task for economists is to study institutions and the processes of institutional conservation, innovation and change.
4. The economy is an open and evolving system, situated in a natural environment, effected by technological changes, and embedded in a broader set of social, cultural, political, and power relationships.
5. The notion of individual agents as utility-maximising is regarded as inadequate or erroneous. Institutionalism does not take the individual as given. Individuals are affected by their institutional and cultural situations. Hence individuals do not simply (intentionally or unintentionally) create institutions. Through "reconstitutive downward causation" [Hodgson 2000] institutions affect individuals in fundamental ways.

Most of these points are direct elaborations of ideas from Hamilton's [1919] text. However, regarding point (4), Hamilton did not mention the words "open system." The phrase did not become widely used until after 1945. Institutional economists such as K. William Kapp [1968, 8] and Shigeto Tsuru [1993, 73] made the idea of the economy as an open system one of the defining characteristics of institutionalism. Furthermore, Hamilton did not use the words "evolving" or "evolutionary," although institutionalists have become fond of these terms.

Point (1) may prove controversial, so it will be discussed in more detail below. It is perhaps the only point that any institutionalist may wish to remove from the list. Certainly, some institutionalists will wish to add to or elaborate on the above five points. The contention here is that they contain the "hard core" of the institutionalist tradition.

I further assert that the single most important defining characteristic of the old institutionalism is proposition (5). Among other schools, the new is distinguished from the old institutional economics principally in these terms. Other criteria do not demarcate the old institutionalism so readily. Other schools of economic thought also express some concordance with propositions (1) to (4). In contrast, proposition (5) is a guiding thread through the whole institutionalist tradition, from Veblen to Galbraith, and it is rarely acknowledged or developed elsewhere. I make this argument below. Let us first look at proposition (1). Subsequently, later sections of this essay will examine the common features of institutionalist theory and discuss some of the implications.

Is Institutionalism Defined by Its Policy Pronouncements?

In the wider world, economics is often perceived and judged in terms of its policy prescriptions. Economics claims to be a science, but policy issues appear everywhere. Even those who adhere to the notion of a "value-free" economic science are often the very same people who are keen to pronounce policies.

The institutional economist Gunnar Myrdal is well known for his emphasis on the unavoidability of value judgments in social science. He wrote: "Valuations are present in our problems even if we pretend to expel them. The attempt to eradicate biases by trying to keep out the valuations themselves is a hopeless and misguided venture" [Myrdal 1958, 131]. But this does not mean that positive and normative statements are epistemologically indistinguishable. For Myrdal, facts and values were not the same thing. Values neither "emerge automatically" from facts nor is the choice of value premises an arbitrary matter. In short, Myrdal believed that "values are always with us," but he did not make the mistake of treating values as equivalent to facts. In social science, statements about fact are always contaminated with values. However, this does not mean that facts and values are equivalent.

Economic policies are very important. Nevertheless, to convince and carry scientific authority, policies have to claim a theoretical basis. Whether from the political right or left, a policy in the modern world has to invoke to some theoretical justification. For reasons of both legitimization and logic, policy has to attempt to ground itself upon theory. Furthermore, in order to change the world it is first necessary to understand it. We must discern its underlying structures and forces before we can appraise the set of feasible possibilities for policy.

It is not being suggested that the positive *and* the normative can be entirely separated, at least in the social sciences. Contrary to the "positive economics" proposed in the neoclassical textbooks, it is impossible to separate completely (positive) judgments of fact from (normative) judgments of value. Statements of fact and value are typically intermixed. However, facts and values are not the same thing.

Accepting that normative values are "always with us" does not mean that we should always judge a theory primarily by its normative values. To accept a complex interrelationship between the positive and the normative does not mean that we abandon all aspects of the distinction. Statements attempting to explain what *is* are confused with statements about what *ought* to be. Yet knowing that many people in the world today are poor is not the same thing as saying that they should remain impoverished.

While important, normative aspects of institutionalism are not very useful, nor sufficiently precise, as defining criteria. One can find a huge diversity of normative opinions within institutionalism. There are prominent examples of fairly conservative institutionalists, such as Arthur F. Burns – a friend of, and collaborator with, Veblen's student Wesley Mitchell – who advised Republican President Eisenhower in the 1950s. Other institutionalists have socialist views. Others are closer to the political center. Policy outputs do not tell us very much about the overall nature of institutionalism.

Any attempt to define institutionalism in terms of policy outputs would run into severe difficulties. Consider some possible policies. Can institutionalism be defined, in part, in terms of a critique of market solutions to economic problems?

Many institutionalists have criticized pro-market policies and have proposed various forms of economic intervention and planning. However, so too have neoclassical economists. (Neoclassical economics is defined as the type of economics invoking the standard textbook principles of rationality, maximization and equilibrium.) The problem of using a disposition towards planning and against markets to define institutionalism would be that many neoclassical economists would then be institutionalists.

Many of the pioneers of neoclassical economic theory, including Léon Walras, Alfred Marshall and Philip Wicksteed, were sympathetic to socialist or social-democratic ideas. By today's standards, some of them would be leftist radicals. Walras, for instance, called himself a "scientific socialist." His theoretical efforts in economics were motivated by a desire to demonstrate the economic advantages of price regulation and the public ownership of natural monopolies, including land. Marshall was concerned about the problems of poverty in Victorian Britain, and was sympathetic to worker co-operatives. Wicksteed also advocated land nationalization and had sympathetic and personal links with the socialist and radical movement.

Several neoclassical economists have promoted radical, interventionist or socialist ideas. For example, Irving Fisher advocated substantial reflationary measures during the Great Depression. Another group of neoclassical economists in the 1930s–led by Oskar Lange–used neoclassical economic tools to argue for the superiority of a version of socialist planning.

Still later, leading neoclassical general equilibrium theorists Kenneth Arrow and Frank Hahn declared their sympathies for various interventionist and social-democratic economic policies. Indeed, Hahn and others have attempted to justify the whole general equilibrium theoretical project as an attempt to demonstrate the *limits* of the market mechanism.

Even more recently, alleged Marxists such as Jon Elster and John Roemer have explicitly embraced neoclassical tools of economic analysis, while retaining leftist political credentials. True, there are many conservative and pro-market neoclassical economists. But neoclassical theory spans the conventional political spectrum – from the extreme pro-planning left to the extreme pro-market right – and thus is not definable in terms of the policy stances of its adherents.

Can institutionalism be defined, in part, in terms of a concern for greater equality and wealth? Institutionalists do not have a monopoly on egalitarian sentiments. And there is nothing in the core of neoclassical theory that necessarily leads us to inegalitarian conclusions.

Indeed, in the early part of the twentieth century, some economists saw neoclassical utility theory as supporting the policy prescription of income redistribution and greater equality. If individuals have a diminishing marginal cardinal utility of income, then making incomes more equal may increase overall utility. However, these egalitarian policies did not find ideological favor among many neoclassical economists; they adopted the Pareto criterion instead. A policy of taking from the rich and giving to the poor is not Pareto efficient. With this auxiliary assumption, the policy conclusions of neoclassical welfare theory were changed from egalitarian to conservative. The core presuppositions of neoclassical theory are in fact enormously flexible in policy terms, depending upon which auxiliary assumptions are chosen.

True, neoclassical theory is based on the idea of the given individual. And the ideology of political individualism sits quite comfortably upon it. But the assumption of given, utility-maximising individuals does not itself contain any normative notion concerning the maximization of human freedom or the minimization of the role of the state. It is one thing to say that the analytical and the normative ideas may dovetail easily. But this does not necessarily mean that one flows logically from the other.

The fact that neoclassical theory can readily be packaged as either pro-market or anti-market is a symptom of its failure to provide an adequate explanation of how markets work. It really concedes too much to neoclassical theory to suggest that it has an adequate theoretical foundation upon which to build any pro- (or anti-) market policy. Neoclassical theory is essentially neither pro-market nor anti-market, because it has no adequate theory of markets at all. Instead of associating it with markets, it would be more accurate to say that neoclassical theory was *blind* to real markets, and consequently to their virtues or vices.

It is a serious mistake to dismiss mainstream economics on policy grounds, especially if we are concerned about policy. The mistake becomes more serious because it gives unwarranted credence to mainstream theory as a means to generate well-grounded policies. The dismissal of a theory because of its alleged policies unwittingly bolsters the theory, by giving it much more credit, as a viable policy engine, than it deserves.

Furthermore, turning science into an ideology would disable any attempt to get better scientific explanations of social and economic outcomes we may wish to change. Instead of persuading the scientific community of the causes of poverty or unemployment, we simply take up an ideological posture against it. We thereby abandon our role and duty as scientists. Our ability to change and improve the world is diminished by some degree. Any alternative approach to the mainstream must first stake its claim to be an identifiable approach to economics on the basis of its incisive *analysis of what is*, rather than on its judgments of what *ought* to be.

Other Criteria: Interdisciplinarity, Institutions, Evolution, and Open Systems

We now consider three more of the defining characteristics of institutionalism, as listed above, namely (2), (3) and (4). I argue that these are necessary, but far from sufficient, to define institutionalism.

Consider the worthy attribute of interdisciplinarity. It is to its merit and enrichment that the old institutional economics draws upon other disciplines such as anthropology, sociology, political science, and psychology.

However, the nature of interdisciplinarity is difficult to pin down. Neoclassical economics could also claim to draw on other disciplines. Chicago economists Gary Becker and Jack Hirshleifer have asserted that their economics makes use of insights from biology. Political science and sociology have been invaded by neoclassical approaches based on rational choice. The neoclassical economist can also be comfortable with some individualistic schools of thought in anthropology and psychology.

Furthermore, not all interdisciplinary endeavors are worthwhile. Many disciplines contin individualist and other assumptions from which institutionalism would disassociate. A richer concept of the individual may also be found in anthropology or psychology, but we also find impoverished and unsuitable ideas in these disciplines. Institutionalists may be more thorough and committed in their use of interdisciplinary resources, but interdisciplinarity does not define institutionalism.

The old institutionalism emphasizes the importance of institutions in economic life, and attempts to understand their role and their evolution. Especially in the 1940–1975 period, mainstream economists neglected the study of institutions. This is not the case today. With the arrival of the new institutional economics, mainstream economists have analyzed institutions, albeit as outcomes of decisions of rational, maximizing agents. The old institutionalists, cannot claim to be the only school of economics to study institutions.

Consider the idea that institutional economics is "evolutionary." The nugget of truth here is that institutionalist writing is concerned with processes of structural transformation, emergence and change, which are often neglected in the mainstream literature. The problem, however, is that the word "evolutionary" is extremely vague. It is now widely used, even by economists using neoclassical techniques. "Evolutionary game theory" is highly fashionable. Even Walras is described as an evolutionary economist [Jolink 1996]. Above all, "evolutionary" is now a voguish word that everyone seems keen to use. In precise terms it signifies little or nothing. Some take it to mean the use of biological analogies; other self-proclaimed "evolutionary economists" see no value in them. A narrower and more precise meaning of "evolutionary" that successfully demarcates institutionalism from other approaches has not yet been elucidated or adopted [Hodgson 1993b, 1999].

We come to the institutionalist understanding of the economy as an "open system." This is clearly an important insight of the old institutional economics, at least in the sense that it is recognized that the economy is part of a natural environment, embodied in a system of social relations, and affected by technological and other changes. So far so good. The problem in using this as a demarcation criterion is that more substance needs to be given to the notion of a "system," and more explanation is required of the characteristic of it being "open" as opposed to "closed." The idea of a system is an important but difficult concept. It connotes some idea of a closely structured interaction between interdependent components. But the boundary of the system may be fuzzy and difficult to establish.

What is an *open* system? Arguably, it is a system that is open to flows of matter, energy or information across its boundary – a system in actual or potential interaction with its environment. Is a national economy engaging in trade with other countries an open system? If so, then standard neoclassical macroeconomics has also embraced open systems. Insofar as neoclassical economics deals with the environmental impact of economic activity, it might also be said to be dealing with an open system.

A narrow version of the "open system" doctrine could rule out a significant fraction of the institutionalist literature, whereas a wider version would also admit much of neoclassical theory. The open system doctrine is not a precise signifier of the historical boundaries of institutionalism. Until it receives further refinement, it is at best an important but imperfect criterion.

In summary, the first four characteristics, (1) to (4), are important but not sufficient to define the old institutionalism. Taken separately, or together in any combination, they are not enough. We must turn to the fifth criterion.

The Institutionalized Individual

The first task in this section is to identify a common theme that pervades institutionalism, from the writings of Veblen in the 1890s and after, to Galbraith and the present day. A notion that the individual is not given, but can be reconstituted by institutions, pervades the tradition of old institutionalism from its predecessors in the historical school to its modern successors. For instance, Veblen [1899, 190–1] wrote:

> The situation of today shapes the institutions of tomorrow through a selective, coercive process, by acting upon men's habitual view of things, and so altering or fortifying a point of view or a mental attitude handed down from the past.

For Veblen, this was a basis for a fundamental critique of mainstream economics. In 1909, he elaborated the argument more fully:

> The wants and desires, the end and the aim, the ways and the means, the amplitude and drift of the individual's conduct are functions of an institutional variable that is of a highly complex and wholly unstable character [Veblen 1919, 242–3].

Likewise, Hamilton [1919, 318] wrote of the "most important" defect of neoclassical economics: "it neglected the influence exercised over conduct by the scheme of institutions under which one lives and must seek his good." Later he continued the same theme, seeing each institution as "imposing its pattern of conduct upon the activities of men" in a manner consistent with the notion that institutions possess causal powers above that of individuals alone. Hamilton [1932, 89] continued: "Institutions and human actions, complements and antitheses, are forever remaking each other in the endless drams of the social process."

Writing in 1899, Commons [1965, 3] saw institutions "shaping each individual." Commons [1934, 73–4] likewise made it clear that "the individual with whom we are dealing is the Institutionalized Mind.... Individuals ... meet each other ... prepared more or less by habit, induced by the pressure of custom ... "

In an early article, Mitchell [1910, 203] made a similar point:

> Social concepts are the core of social institutions. The latter are but prevalent habits of thought which have gained general acceptance as norms for guiding conduct. In this form the social concepts attain a certain prescriptive authority over the individual. The daily use by all members of a social group unremittingly molds those individuals into common patterns without their knowledge, and occasionally interposes definite obstacles in the path of men who wish to act in original ways.

In his study of the evolution of money as an institution, Mitchell [1937, 371] emphasized how it changed human mentality and nature:

> Now the money economy ... is in fact one of the most potent institutions in our whole culture. In sober truth it stamps its pattern upon wayward human nature, makes us all react in standard ways to the standard stimuli it offers, and affects our very ideals of what is good, beautiful and true.

Likewise, Clarence Ayres [1944, 84] explained:

> "wants" are not primary. They are not inborn physical mechanisms and they are certainly not spiritual attributes. They are social habits. For every individual their point of origin is in the mores of his community; and even these traditions have a natural history and are subject to modification in the general process of social change.

The idea that individual tastes are not given, but are shaped by institutional circumstances and by particular influences such as advertising, is a major theme in the writings of Galbraith. For instance, in the *New Industrial State*, Galbraith [1969, 152] insisted that individual "wants can be synthesized by advertising, catalysed by salesmanship, and shaped by the discreet manipulations of the persuaders." The theme persists throughout his writings. Indeed, no author has brought these ideas to the attention of the modern reader more clearly and resolutely than Galbraith. His analysis puts particular emphasis on the effects of advertising on individual wants. This is one version of the core institutionalist story. More generally, institutionalists recognize the potential influence of many institutions on individual habits, conceptions, and preferences.

Such ideas permeate and endure through institutionalism as a whole. Institutionalism is distinguished from both mainstream economics and the "new institutional economics" precisely for the reason that it does not assume a given individual, with given purposes or preference functions.

Instead of a bedrock of given individual, presumed by the mainstream and new institutional economics, the old institutionalism holds to the idea of interactive and partially malleable agents, mutually entwined in a web of partially durable and self-reinforcing institutions [Hodgson 1988]. No other criterion demarcates so clearly the old institutional economics, on the one hand, from new institutional and mainstream economics on the other [Hodgson 1993a].

Note that the acceptance of the institutionalized individual does not immediately rule out the possibility that institutionalism and neoclassical economics may be complementary. Although Veblen wished to purge economics of classical and neoclassical errors, other institutionalists searched for some complementarity between neoclassical and institutional economics. This group included leading institutionalists such as Commons, Mitchell, J. M. Clark, Paul Douglas, and Arthur F. Burns. They all saw institutionalism as compatible with aspects of Marshallian price theory. Commons [1931] in particular argued for some complementarity between the schools. This is a controversial position. But it shows that the complete exclusion of neoclassical economics from institutionalism would rule out Commons and others from the institutionalist canon.

Upward and Downward Causation

Having identified the most important common theme in old institutionalism, it is necessary to enquire more deeply into its meaning. Several versions of this doctrine have surfaced over the years. It is also necessary to deal with some potential misunderstandings and rebuttals.

Perhaps the most frequent attack on the notion that individual tastes and preferences are molded by circumstances is the criticism that this leads to some kind of structural or cultural determinism. The individual, it is said, is made a puppet of social or cultural circumstances.

Admittedly, some old institutionalists have promoted such a deterministic view. When Ayres [1961, 175] wrote that "there is no such thing as an individual" he was giving succor to such ideas [Rutherford 1994, 40–41]. The danger is to see social order as a primarily "top down" process in which individuals are formed and cajoled by institutions, with a neglect of individual autonomy and agency. The Ayresian version of institutionalism has been so prominent in the post-1945 era that many commentators take it to be representative of institutionalism as a whole.

However, such predominantly "top down" versions of the core institutionalist idea are not common to all institutionalists. This is clearly the case with both Veblen and Commons. For instance, Veblen [1919, 243] argues that institutions are the outcome of individual behavior and habituation, as well as institutions affecting individuals:

> The growth and mutations of the institutional fabric are an outcome of the conduct of the individual members of the group, since it is out of the experience of the individuals, through the habituation of individuals, that institutions arise; and it is in this same experience that these institutions act to direct and define the aims and end of conduct.

Writing in 1899, Commons [1965, 6–8] wrote similarly of the dependence of institutions upon beliefs:

> Social beliefs . . . furnish the basis in the affections of each person which alone makes possible his responsiveness to the appeals of those with whom he must coöperate. The institution in which he finds himself is both the cause and effect of his beliefs. . . . Common beliefs and desires are the vitalizing, active force within the institution.

These statements show a valid recognition of both the dependence of institutions upon individuals and the molding of individuals by institutions. In the writings of Veblen and Commons there is both upward and downward causation; individuals create and change institutions, just as institutions mold and constrain individuals. Institutionalism is not necessarily confined to the "top down" cultural and institutional determinism with which it is sometimes associated.

The great merit of the institutionalist idea that institutions shape individual behavior is that it admits an enhanced concept of power into economic analysis. Power is not simply coercion. For Steven Lukes [1974], the overemphasis on the coercive aspect of power ignores the way that it is often exercised more subtly – and often without overt conflict. He points out that supreme power is exercised by orchestrating the thoughts and desires of others.

These subtle considerations are absent from mainstream economics. Preference functions are not subject to reconstitutive downward causation. This is so even when an attempt is made to "explain" tastes. Becker [1996] tries to show that cultural and other influences can alter preferred outcomes by adding cultural and other factors to the arguments of these functions. However, culture does not alter the preference functions themselves.

A problem with this analysis is that it cannot deal with the genuine evolution and fundamental development of the individual. It attempts to make all explanations of social phenomena reducible to the given individual, but in doing so it has to make the individual preference function immutable. The preference function is already "there," ready to deal with unpredictable and unknowable circumstances. We already know essentially what is to be learned.

Learning typically takes place through and within social structures, and at least in this sense it is an important case of reconstitutive downward causation. Neoclassical economics has great difficulty accommodating the notion of learning because the very idea of "rational learning" is problematic. It treats learning as the cumulative discovery of pre-existing "blueprint" information, as stimulus and response, or as the Bayesian updating of subjective probability estimates in the light of incoming data. However, instead of the mere input of "facts" to given individuals, learning is a developmental and reconstitutive process. Learning involves adaptation to changing circumstances, and such adaptations mean the reconstitution of the individuals involved. Furthermore, institutions and cultures play a vital role in establishing the concepts and norms of the learning process [Hodgson 1988]. Accordingly, the reconstitutive nature of learning is partly a matter of reconstitutive downward causation. To put it bluntly: if we are to accept fully the notion of learning into social theory, then the concept of reconstitutive downward causation must also be sanctioned.

Conclusion

It has been argued here that a concern for policy issues may be an attribute of institutional economics but it cannot be its defining characteristic. Necessary features of institutionalism include the recognition of the importance of insights from other disciplines, of institutions and of open and evolving systems.

Nevertheless, the single most important characteristic of institutionalism is the idea that the individual is socially and institutionally constituted. The argument here is that all the old institutional economists, from Veblen to Galbraith, embrace the notion that the individual is molded by cultural or institutional circumstances. Within institutionalism, there are many variants of this view.

However, the notion of "reconstitutive downward causation" is found neither in mainstream economics, nor in the new institutionalism. The tradition there is to take the individual as given. His or her preference function, even if it includes cultural variables as arguments [Becker 1996], is immanently conceived. The emphasis there is on "rational choice" with given preferences in specified circumstances. Welfare judgments are based on the assumption that the choice made by the individual is the "best" one in the circumstances.

The implications for abandoning this view and adopting the approach of the old institutionalism are enormous. Conceptions of social power and learning can be placed at the center of economic analysis. This means that institutionalism is more able to address questions of structural change and economic development. It is more useful, for instance, in dealing with issues such as long-term economic development, the problems of less-developed economies, or the transformation processes in the former Soviet bloc countries. On the other hand, the analysis becomes much more complicated and less open to formal modelling. In normative terms, the individual is no longer taken as the best judge of his or her welfare. This opens up the difficult question of the discernment and evaluation of human needs.

This theoretical agenda – including matters of power, learning and welfare – is at the center of institutionalism, and it remains as vital and exciting as it was 100 years ago.

References

Ayres, Clarence E. *The Theory of Economic Progress.* 1st ed. Chapel Hill, N.C.: University of North Carolina Press, 1944.

______. *Toward a Reasonable Society: The Values of Industrial Civilization.* Austin: University of Texas Press, 1961.

Becker, Gary S. *Accounting for Tastes.* Cambridge, Mass.: Harvard University Press, 1996.

Commons, John R. "Institutional Economics." *American Economic Review* 21, no. 4 (December 1931): 648–57.

______. *Institutional Economics – Its Place in Political Economy.* New York: Macmillan, 1934.

______. *A Sociological View of Sovereignty*, edited with an introduction by Joseph Dorfman. New York: Augustus Kelley, 1965.

Galbraith, John Kenneth. *The New Industrial State.* Harmondsworth: Penguin, 1969.

Hamilton, Walton H. "The Institutional Approach to Economic Theory." *American Economic Review* 9, Supplement (1919): 309–18.

______. "Institution." In *Encyclopaedia of the Social Sciences*, vol. 8, edited by E. R. A. Seligman and A. Johnson, 84–89. New York: Macmillan, 1932.

Hodgson, Geoffrey M. *Economics and Institutions: A Manifesto for a Modern Institutional Economics.* Cambridge and Philadelphia: Polity Press and University of Pennsylvania Press, 1988.

______. "Institutional Economics: Surveying the 'Old' and the 'New'." *Metroeconomica* 44, no. 1 (February 1993a): 1–28.

______. *Economics and Evolution: Bringing Life Back Into Economics.* Cambridge, U.K., and Ann Arbor, Mich.: Polity Press and University of Michigan Press, 1993b.

______. *Evolution and Institutions: On Evolutionary Economics and the Evolution of Economics.* Cheltenham: Edward Elgar, 1999.

______. "Structures and Institutions: Reflections on Institutionalism, Structuration Theory and Critical Realism." Unpublished mimeo, University of Hertfordshire, 2000.

Jolink, Albert. *The Evolutionist Economics of Léon Walras*. London and New York: Routledge, 1996.
Kapp, K. William. "In Defense of Institutional Economics." *Swedish Journal of Economics* 70 (1968): 1–18.
Lukes, Steven. *Power: A Radical View*. London: Macmillan, 1974.
Mitchell, Wesley C. "The Rationality of Economic Activity. Part II." *Journal of Political Economy* 18, no. 3 (March 1910): 197–216.
———. *The Backward Art of Spending Money and Other Essays*. New York: McGraw-Hill, 1937.
Myrdal, Gunnar. *Value in Social Theory*. New York: Harper, 1958.
Rutherford, Malcolm H. *Institutions in Economics: The Old and the New Institutionalism*. Cambridge: Cambridge University Press, 1994.
Tsuru, Shigeto. *Institutional Economics Revisited*. Cambridge: Cambridge University Press, 1993.
Veblen, Thorstein B. *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*. New York: Macmillan, 1899.
———. *The Place of Science in Modem Civilisation and Other Essays*. New York: Huebsch, 1919.

PART FIVE

NEW DIRECTIONS IN ECONOMIC METHODOLOGY

As explained in the introduction to this volume, there has been a flood of recent and valuable work on economic methodology. Part V provides a tiny selection of this work. In choosing these five articles from hundreds of possibilities, I have attempted both to provide some sense of the range of contemporary works and to sample from some of the most influential methodological approaches and authors.

A brilliant writer, a distinguished economic historian, and a bold innovator, Deirdre N. McCloskey in Chapter 22 challenges the pretense of methodologists to guide practice in economics and argues that rhetoric (the study of persuasion) is a more fruitful and enlightening guide.

Although committed to methodology, Uskali Mäki's and Tony Lawson's realist programs as sketched in Chapters 23 and 24 call for drastic shifts in the way that methodology is done. Both argue that methodologists should be concerned with the ontological commitments of economists – that is with an exploration of what economists take to be real. However Mäki's and Lawson's explorations take very different directions, with Lawson arguing that ontological inquiry leads to serious criticisms of mainstream economics.

In Chapter 25, Julie A. Nelson explores gendered presuppositions within mainstream economics and ways in which sensitivity both to those presuppositions and to the specific economic circumstances of women could improve the discipline. In the last chapter in the volume, Robert Sugden offers a novel account of theoretical models and clues concerning why theoretical models occupy such a dominating role within economics.

My hope is that the essays in this section and indeed in the anthology as a whole will whet the reader's appetite for more.

TWENTY-TWO

The Rhetoric of This Economics

Deirdre N. McCloskey

Deirdre N. McCloskey (1942–) was educated at Harvard University, taught for many years at the University of Chicago and the University of Iowa, and is currently a UIC Distinguished Professor of Economics, History, English, and Communication at the University of Illinois at Chicago. McCloskey was Donald until 1995. She describes her transition in *Crossing: A Memoir*. In addition to her long-standing research interests in economic history, in the 1980s McCloskey became interested in the ways in which economists persuade one another, and her work on the rhetoric of economics poses a serious challenge to traditional views of economic methodology. The author of twenty books and three hundred articles, McCloskey has been an extremely influential figure.

In the opening scene of the movie *The Graduate* a Mr. McGuire puts an avuncular arm around the Dustin Hoffman character and says, "I just want to say one word to you. Just one word." Yes, sir? "Are you listening?" Yes, I am. "Plastics." [Pause] Exactly how do you mean it? "There's a great future in plastics. Think about it. Will you think about it?" Yes, I will. "Enough said: that's a deal."

So nowadays the avuncular word to the wise is "rhetoric." There's a great future in rhetoric. Furthermore, unlike plastics, rhetoric has also had a great past, the twenty centuries during which it was the educator of the young and the theory of speech in the West – as the classicist Werner Jaeger called it, "the first humanism," the "rhetorical paideia." The three and a half centuries of modernity since Bacon and Descartes have been in this respect an interlude. "We are still bemused," notes Richard Lanham the historian of rhetoric, "by the 300 years of Great Newtonian Simplification which made 'rhetoric' a

Pages 38–52 of *Knowledge and Persuasion in Economics* by Deirdre N. McCloskey. Near the end of the essay, three paragraphs concerning a no-longer current political issue were deleted with the author's permission.

dirty word, but we are beginning to outgrow it" (1993, ch. 2, p. 27; cf. Lanham 1992). British empiricism and French rationalism have had a long and glorious run. The revival of rhetoric has been explicit since the 1960s in the study of literature and speech. But a sense of how to do things with words has spread now to other inquiries, to philosophers ruminating on speech acts or linguists on the pragmatics of conversation.

Rhetoric in the late twentieth century has had to be reinvented in ignorance of its past. Yet the mathematician who reflects on the standard of proof in topology or the economist who notes that the Federal Reserve Board is a speaker with intent or the political scientist who wonders amidst his regression equations if politics should after all be reduced to public opinion polls (Barry 1965; J. Nelson 1983) are practicing rhetoric. When they reflect on their reflections they are practicing, to say just three words to you again – are you listening? – the "rhetoric of inquiry."

When Kenneth Arrow was asked by George Feiwel what criteria he uses to judge competing theories in economics he answered:

> Persuasiveness. Does it correspond to our understanding of the economic world? I think it foolish to say that we rely on hard empirical evidence completely. A very important part of it is just our perception of the economic world. If you find a new concept, the question is, does it illuminate your perception? Do you feel you understand what is going on in everyday life? Of course, whether it fits empirical and other tests is also important. (Feiwel 1987, p. 242)

Surprisingly the passage is quoted by Mark Blaug as demonstrating that Arrow is a Lakatosian (Blaug 1991, p. 505). Its prose meaning, though, is that Arrow, like us all, is a rhetorician. He seeks persuasion, through introspection, through a sense of the social world, and through fully identified best linear unbiased econometric tests, too.

The very word "rhetoric," though, makes it hard for moderns to understand what they are talking about. Like "anarchism" taken to be bomb-throwing or "pragmatism" taken to be unprincipled horse trading, rhetoric is a noble word fallen on bad times.

Rhetoric has since the beginning been defined in two ways, as I have said, one narrow and the other broad. The narrow definition is Plato's, made popular in the nineteenth century by the Romantic elevation of sincerity to the chief virtue. "Rhetoric" in the Platonic definition is cosmetic, hiding a disease under paint rather than providing a cure. Journalists use the cosmetic definition in their news stories and philosophers use it in their seminars. When the newspapers want to speak of obscuring blather and thirty-second spots on flag burning they write "Senate Campaign Mired in

Rhetoric." The philosophy seminar uses the word "rhetoric" to characterize the meretricious ornament obscuring the clear and distinct idea. Thus even W. V. Quine, in an untutored entry for "Rhetoric" in his personal dictionary of philosophy, calls it "the rallying point for advertisers, trial lawyers, politicians, and debating teams" (Quine 1987, p. 183), without noticing that even in such a sneering and Platonic definition it is the rallying point also for philosophers.

In Plato's language "rhetoric" is associated especially with democratic institutions, such as assemblies or law courts, disdained by men of taste. "You attempt to refute me," says Socrates in the *Gorgias,* "in a rhetorical fashion, as they understand refuting in the law courts . . . But this sort of refutation is quite useless for getting at the truth." Or in the *Phaedrus,* "he who is to be a competent rhetorician need have nothing at all to do, they say, with truth in considering things which are just or good, or men who are so, whether by nature or by education. For in the courts, they say, nobody cares for truth about these matters, but for what is convincing" (*Gorgias*, 471e and *Phaedrus* 272d). Compare *Gorgias* 473e-474a: "Polus, I am not one of your statesmen . . . The many I dismiss" (cf. 471e, 502e on rhetoric as mere flattery); and *Phaedrus,* 260a, 275e, 277e, 267a-b, 261c-d, 262c, among other places where Plato expresses his contempt for law courts and democratic assemblies as against those who know. The attack on rhetoric has more than a little anti-democratic coloring.

If rhetoric is defined thus as ornament it is easily left to the "goddamned English professors" or advertising flacks. The setting aside began with Peter Ramus in the sixteenth century, who disastrously reaffirmed the Platonic separation of mere ornament from deep philosophy. As Lanham notes, "If you separate the discipline of discourse into essence and ornament, into philosophy and rhetoric, and make each a separate discipline, it makes them easier to think about. Thus begins modern inquiry's long history of looking for its lost keys not where it lost them but under the lamppost, where they are easier to find" (ch. 7, pp. 6–7). Another professor of English has warned against sneering at the "mere" rhetoric: we must "ward off the sensation that words are nothing but words when they are actually among our most substantial collective realities" (Petrey 1990, p. 37). Our politics, for example, is a set of speech acts and speeches about speech acts, and is easily corrupted by bad rhetoric. "We are only men," wrote Montaigne, "and we only hold one to the other by our word." (I:9).

The other, broad definition of rhetoric is Aristotle's, in *The Rhetoric,* I. II. 1 (to quote the Kennedy translation), "an ability, in each [particular case], to see the available means of persuasion." Of course the Greeks, ever talkers and fighters, distinguished sharply between persuasion (*peitho*) and violence (*bia*), an opposition finely discussed by Kirby (1990). Their literature is filled with speeches of persuasion weighing against the violent alternative. King Priam of Troy, prostrate before Achilles, pleads eloquently for the body of his son, linking in his final words the instruments of persuasion and of violence: "I put my lips to the hands of the man who has killed my children" (Homer, *Iliad* XXIV, line 506). The Athenians at the height of success in the Peloponnesian War sneer at "a great mass of words that nobody would believe," mere rhetoric. They tell the Melians, their victims, that as a matter of realism in foreign policy – compare the rhetoric of Henry Kissinger and the 1960s movement to "realism" in international relations – "the standard of justice depends on the equality of power to compel" (Thucydides, V, 89). The Athenians proceed to kill all the men and sell the women and children into slavery, an abandonment of sweet persuasion they live to regret.

All that moves without violence, then, is persuasion, *peitho,* the realm of rhetoric, unforced agreement, mutually advantageous intellectual exchange. It would therefore include logic and fact as much as metaphor and story. "Logic," as logicians have been making steadily clearer in the century past, is not an unargued realm. Logic can be Aristotelian, scholastic, first-order predicate, deontic, modal, relevant, multivalued, informal, intensional, counterfactual, epistemic, paraconsistent, relevant entailment, fuzzy, and so on and so forth through the various ways that people can formalize what they are saying. The linguist and logician James D. McCawley says that "only through arrogance or ignorance do logicians palm off any single full system of logic as unchallengeable" (McCawley 1990, p. 378). Likewise "fact" is not to be determined merely by kicking stones or knocking tables. That a fact is a fact relative only to a conceptual scheme is no longer controversial, if it ever was. Kant knew it; so should we. Studies of science over the past few decades have shown repeatedly that facts are constructed by words.

There is nothing shameful in this logic and fact of scientific rhetoric. As Niels Bohr said, "It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature . . . We are suspended in language . . . The word 'reality' is also a word, a word which we must learn to use correctly" (Moore 1966 [1985], p. 406; but not all people are gifted at every part of argument; Bohr, gifted at metaphor, could not follow the plots of his beloved movie westerns, and would bring someone along to whisper explanations in his ear). And Heisenberg: "Natural science does not simply describe and explain nature; it is part of the interplay between nature and ourselves: it describes nature as exposed to our method of questioning" (1959, quoted in Berger 1985, p. 176). That is to say, appeals to experimental finding are as much a part of a broad-church definition of rhetoric as are appeals to the good character of the speaker. Mill's logic of strict implication is as much rhetoric as is the anaphora of Whitman's poetry. Wittgenstein says, "Uttering a word is like striking a note on the keyboard of the imagination" (1945 [1958], p. 4). In this definition a science as much as a literature has a "rhetoric."

When economists look at something, say childcare, they think of markets. "Childcare" – which to other people looks like a piece of social control, or a set of buildings, or a problem in social work – looks to economists like a stock certificate traded on the New York exchange. By this choice of metaphor they are driven to identify a demand curve, a supply curve, and a price. If the economists are of the mainstream, neoclassical kind they will see "rational" behavior in such a market; if they are Marxist or institutionalist or Austrian economists they will see somewhat differently. But in any case the seeing will seem to them to make ordinary sense, to be the way things really truly are.

A rhetorician, however, notes that the "market" is "just" a figure of speech. Yet a serious rhetorician, or a serious philosopher of science, will not add the "just," because metaphor is a serious figure of argument. Noting the metaphors is not merely another way of saying that economics is approximate and unperfected. Economists believe that metaphor comes from the fuzzy, humanistic side of the modernist world. A model in economics comes to be called a metaphor, in this way of thinking, if "the statement can be tested only approximately" (thus David Gordon 1991). But the inverse square law of gravitational attraction is also a metaphor; so is Einstein's generalization, It is well known that the Romantics assigned metaphor to the realm of art, distinguishing an imaginative from a scientific faculty, as though different organs of the brain. The literary critic Francis McGrath has argued that the distinction cannot be sustained (McGrath 1985). Boyle's Law shares metaphor with Shakespeare's 73rd sonnet: metaphor, McGrath argues, is as fundamental to science as to art.

Models are metaphors, that is all. So in other fields: "the mechanistic, . . . the organismic, the marketplace, the dramaturgical, and the rule-following metaphors have all played a significant role in psychological research of the past decades" (Gergen 1986, p. 146). "The market" is a commonplace, a *locus communis, a topos* – a place where economists work. The rhetorician's metaphor here is locational. In the rhetorical way of talking since the Greeks the metaphor of a "conversation" is a *topos* for the language game across the playing fields of economics (Klamer and Leonard 1993 explore metaphors in economics more thoroughly, with reference to the now-large philosophical literature; and see McCloskey 1985).

The conversational figure of speech suggests the Similarity Argument: that the economic conversation shares many features with other conversations differently placed. Any scientific conversation has much in common with, say, poetic conversation, as is demonstrable in detail beyond rational patience. The linguist Solomon Marcus listed fully fifty-two alleged differences between scientific and poetic communication (rational vs. emotional; explicable vs. ineffable; and so forth), and after much thought rejected them all as crudities (Marcus 1974). He noted that there is as much variation within scientific and poetic communication as between them.

The attempts to distinguish the artistic and scientific uses of metaphor presume that the categories of European thought around 1860 cut the universe at its joints. The English professor Richard Lanham argues at length that "nothing but confusion has ever come from the effort to fix the poetry-prose boundary" (Lanham 1974, p. 65). Attempts to distinguish art and science do not seem to work, though from the best workers. Thomas Kuhn, for example, noting truly that "we have only begun to discover the benefits of seeing science and art as one" (1977, p. 343), nonetheless attempts a distinction. He argues that beauty in science (a differential equation with startlingly simple solutions, say) is an input into the solution of a technical problem, whereas in art the solution of a technical problem (contrapposto in representing a standing figure, say) is an input into the beauty. But at different levels of the art and science different work will be done. An economic scientist will work like an artist at a technical problem to achieve beauty; but then the beauty at another level will work to solve a technical problem. One might stand better amazed, as a physicist famously did of mathematics, about the unreasonable effectiveness of aesthetic standards in science. The physicist Tullio Regge remarked to Primo Levi, the chemist and writer, "I liked the sentence in which you say that the periodic table is poetry, and besides it even rhymes" (Levi and Regge 1992, p. 9). Levi responded, "The expression is paradoxical, but the rhymes are actually there . . . To discern or create a symmetry, 'put something in its proper place,' is a mental adventure common to the poet and the scientist" (pp. 9–10).

The one distinction between art and science of which Kuhn half persuades me is that art continues to converse with dead artists. Physicists, notoriously, do not work in the past of their discipline. And yet: Biologists are still conversing with Darwin, economists with Adam Smith. Even this most persuasive demarcation seems fuzzy and useless. One can ask of the cleverest of demarcation criteria, so what? In many of the activities of artists and scientists you can see and use the overlap. What is the corresponding usefulness of demarcating science from art?

Logic, for example, is by no means the sole preserve of calculators. The English Metaphysical poets of the seventeenth century were addicted to logical forms, forms that were viewed as figures of speech by writers still educated in rhetoric. John Donne's "Song" (1633) begins with a *reductio ad absurdum* ("Go and catch a falling star, / Get with child a mandrake's root / . . . And find / What wind / Serves to advance an honest mind"), turns then to an inferential argument ("Ride ten thousand days and nights . . . And swear / No where / Lives a woman true and fair"), and finishes with what an economist would call an assessment of a low prior probability ("If thou find'st one, let me know; / . . . Yet do not; I would not go, / Though at next door we might meet. / . . . Yet she / Will be / False, ere I come, to two or three").

Marvell's "To His Coy Mistress" (1681) is the type of an argumentative poem. The argument is of course economic: had we but world enough and time, my Lady, I could court you as your value warrants, to satiation; but time is scarce, and life especially; the rate of time discount (as the modern economist would say) is therefore positive; and the optimal consumption plan is therefore to seize the day. Marvell makes his appeal relentlessly and smirkingly: he plays with a convention of rational choice and mocks it, as language games have a tendency to do with themselves. (Irony for this reason is called by the literary critic Kenneth Burke the "trope of tropes".) The economist plays no less within a convention when drawing on inference (N = ten thousand days and nights) or time discount (t = Deserts of vast Eternity), or when making little jokes to other economists about "islands" in the labor market or how the data have been "massaged." The flatfooted among economists and poets lack this sense of irony about arguments. They pen lines like "The coefficient is significant at the .00000001 level" or "I think that I shall never see / A poem lovely as a tree."

Similarity is not identity. Economics may be like poetry in this or that important respect, but plainly it is not the same. At another level, the likenesses between stocks and childcare will allow the topos of The Market to work, but there are differences, too, that will figure sometimes. At still another level, academic poets have different conversations from greeting-card poets. And all poets differ in other ways from economists, however poetic the economists may be.

Economics, for example, is *not* poetry just to the degree that a piece of economics invites what the critic Louise Rosenblatt called an "efferent" reading (from Latin *effero,* "take away") as against an aesthetic reading (1978, pp. 25–28). That is, one expects to "take away" something useful from an article on the New Jersey income-maintenance experiment. The article is not read for itself (though recall Marcus' experiment and take care: some economics is read for the aesthetic pleasure, and could hardly give any other). As Oakeshott put it (1959 [1991], p. 525), "poetic utterance . . . is not the 'expression' of an experience, it *is* the experience and the only one there is" in the voice of poetry.

It is sometimes argued therefore that economics and other sciences, though using metaphors, use them in a different way from poets. The philosopher of science Cristina Bicchieri, for example, in a penetrating comment on my "poetics" of economics, argues that "A good literary metaphor should be surprising and unexpected . . . Scientific metaphors, on the contrary, *are to be overused*" (1988, p. 113, my italics; compare Oakeshott 1959 [1991], p. 528: the poet's "metaphors have no settled value; they have only the value he succeeds in giving them").

Well, yes and no. The economist A. C. Harberger tells the story of a cocktail party at his house in the early 1960s, when Gary Becker, a brilliant student at Chicago, was working on the dissertation that became *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (1964). The party was well along, but Gary as usual was sober and serious, and always, always talking economics. He came up to Harberger and remarked out of the blue, soft drink in hand, "You know, Al, *children are just like consumer durables.*" It was a poetic moment, unexpected certainly to Harberger (who in fact was an expert on consumer durables, but had no idea that procreation might fit the category). True, as Bicchieri says, Becker intended the metaphor "to be overused," which is to say, to become part of the dead metaphors of the field; and it has. But at the moment of creation – like a poem once alive that becomes a cliché – it was anything but dead.

And on the literary side Bicchieri and other philosophers who want to give scientific metaphors a special "cognitive" goal quite separate from poetry are overstating the strangeness of poetry. They are adopting without realizing it a romantic literary criticism that puts the poet outside the routines of conversation, the poem being "the spontaneous overflow of powerful feeling," taking its origin from "emotion recollected in tranquillity." But of course poets, even Wordsworth, in fact talk largely about poetry, quoting each other's metaphors. The coin of poetic tradition is well worn. Some good poems contain clichés like "the coin of poetic tradition is well worn." What makes the poem work as "the activity of being delighted in the entertainment of its own contemplative images" (Oakeshott 1959 [1991], p. 527) is what is done with the clichés (like what I did just now with the cliché of worn coins, or in this sentence with the convention of not referring to one's own clichés in academic prose, or in this clause with that of not engaging in tiresome reflexivity, or in the last clause with that of not disabling the reader's vexation by admitting that it is "tiresome," and so forth). But good science is like that, too. Good science like good poetry can take utterly routine metaphors and, as Harberger is fond of saying, "make them sing." Periods of classicism, in which a poet or scientist seeks originality within settled metaphors, are not non-poetic or non-scientific. Think of Alexander Pope or Lord Kelvin.

Still, to be less provocative, take a conversation more obviously similar to economics, one which is wholly efferent (maybe), economic journalism. Thinking about its metaphors and contrasting them to economics itself still proves useful. Economic journalism is written sometimes by journalists with no academic pretensions, such as Leonard Silk, Robert Samuelson, John Greenwald, Louis Rukeyser, and David Warsh, but also by academic economists gifted in this way, such as Milton Friedman, J. K. Galbraith, and Lester Thurow, or academics-turned-journalists like Peter Passell. The common reader is liable to think that such writings are academic economics "translated" into plain English, in the style of popular science. Without prejudice, they are not. (Which is not to say that economic journalism is easy or that it is inferior to seminar talk: anyone who could imitate the books by the financial journalist who writes under the pseudonym "Adam Smith," for example, would be justly rich; few academic economists are.)

The journalistic conversation runs on particular dramatic conventions, hinging on evil, suspense, and individuality. William Blundell, a feature writer for *The Wall Street Journal,* gives as "the major commandment" for newspaper reporting: "For Pete's sake, make it interesting. Tell me a *story*" (Blundell 1988, p. xii), and uses the old gag about the ideal *Reader's Digest* piece to make the point: "How I Had Carnal Relations with a Bear for the FBI and Found God." In the storied talk that market people use to dignify their work a market is "excited" or "depressed," overrun with bulls or bears, slit with cutthroat competition. "How I Had Business Relations with IBM for the S.E.C. and Found Competitiveness." Businesspeople are portrayed in a story by Samuel Smiles or Louis Rukeyser as pioneers whose courage and creativity extends the frontiers of what is economically possible; or they are portrayed in a story from Lincoln Steffens or Robert Kuttner as the tyrants who oppress the powerless. The "story" is just that: a piece in a newspaper. The black hat appears in it as a foreign country underselling "our" products or "beating us" in productivity. We and they are the heroes and villains, in pervasive sporting and military metaphors. Personalizing images are common, as in the talk of the street.

A masterful example is *The Zero-Sum Solution* (1985), by Lester Thurow, a fine economist and dean of the business school at Massachusetts Institute of Technology. The book is sporting. "To play a competitive game is not to be a winner – every competitive game has its losers – it is only to be given a chance to win . . . Free market battles can be lost as well as won, and the United States is losing them on world markets" (Thurow 1985, p. 59). One chapter is entitled "Constructing an Efficient Team." Throughout there is talk about America "competing", and "beating" the rest of the world with a "world-class economy." A later book is called *Head to Head.* Thurow complains that more people don't appreciate his favorite metaphor: "For a society which loves team sports . . . it is surprising that Americans won't recognize the same reality in the far more important international economic game" (1985, p. 107). Note that my "reality" is your "metaphor." In more aggressive moods Thurow slips from sweatpants into combat fatigues: "American firms will occasionally be defeated at home and will have no compensating foreign victories" (Thurow 1985, p. 105). Foreign trade is viewed as the economic equivalent of war.

Three metaphors govern Thurow's story: this metaphor of the international zero sum "game"; a metaphor of the domestic "problem"; and a metaphor of "we". *We* have a domestic *problem* of productivity that leads to a loss in the international *game.* Thurow has spent a long time interpreting the world with these linked metaphors. The we-problem-game metaphors are not the usual ones in economics. Anti-economists since the beginning have favored the metaphor of exchange as a zero-sum game. But the subject is the exchange of goods and services. If exchange is a "game" it might better be seen as one in which everyone wins, like aerobic dancing. No problem. Trade in the mainstream economic view is not zero sum. To be sure, from the factory floor it looks like zero sum, which gives Thurow's metaphor the appearance of common sense. To a businessperson "fighting" Japanese competition in making automobiles, her loss is indeed Toyota's gain. But the competitive metaphor looks at only one side of the trade, the selling side. Economists see around and underneath the economy. Underneath it all (as the economists say, in their favorite metaphor) Jim Bourbon of Iowa trades with Tatsuro Saki of Tokyo. A Toyota sold by Japan pays for 2,000 tons of soybeans sold by the United States. But at the same time a Japanese and an American consumer are gaining soybeans and an auto. One kid gets the other kid's pet frog in exchange for giving up his jackknife. Both kids are better off. If we look on nations in the way we look on kids making such exchanges we can see that both nations win a little something.

Trade and development are in the economic metaphor positive sum, not zero sum. The economic metaphor suggests a different attitude towards trade than that of Friedrich List, the German theorist of the German customs union in the early nineteenth century, or Henry Carey, the nineteenth-century American theorist of protection, or Lester Thurow and other recent Jeremiahs of American decline.

Talking in such a rhetorically self-conscious way about a piece of economic journalism is not just a rhetorical trick for attacking it. The point is that all conversations are rhetorical, as I have said, that none can claim to be the Archimedean point from which others can be levered once and for all. The neoclassical economists who would disagree with Thurow, such as his colleague at Massachusetts Institute of Technology, Paul Krugman, use metaphors, too, of humans as calculating machines and rational choosers. The neoclassicals say that the human situation is rational choice, the maximization of an objective function subject to constraints. Their metaphor is less thrilling perhaps than the economy as a struggle between good and evil or as the final round of the National Basketball Association playoffs; but it is no less metaphorical on that count. The rational-choice model is the master metaphor of mainstream economics, enticing one to think "as if" people really made decisions in this way. The metaphor has disciplined the conversation among neoclassical economists – the discipline is: if you don't use it, I won't listen – and has produced much good. To it we owe insights into subjects ranging from the consumption function in the twentieth century to the enclosure movement in the eighteenth. Yet, to repeat, it is a metaphor.

The neoclassicals (I am one of them) are very fond of their metaphor of people as calculating machines. What is problematical is the "positive" and "objective" status they ascribe to it. It was not always so. Ambiguity and contention surrounded the triumph of calculating choice as the definition of economics, as did the triumph of the computer analogy in psychology, and it was by no means always regarded as an innocent analytic technique. More than a century ago William Stanley Jevons found the calculating machine persuasive on the non-positivist grounds that it fitted with Bentham's calculus of pleasure and pain; Vilfredo Pareto, too, credited it in the early years of this century with psychological significance.

The neoclassical conversation about the logic of choice, despite the centripetal force of a mathematics teachable to all, has itself tended to break into smaller groups. The new classical macroeconomist has enchanted many young economists, with their lust for certitude. The neo-Keynesian, once himself lusty, holds back, finding solace in tales of Akerlof and sayings of Sen (Klamer 1983, 1984). The other heirs of Adam Smith diverge more sharply from the faith. Even when educated in neoclassical economics, for example, the Marxist economist will object to the neoclassical reduction of the social to the individual; the Austrian economist will object on the other hand to the aggregation of the individual in the social. The Marxist prefers a conversation about the class basis of work; the Austrian prefers a conversation about the ineffable individuality of the entrepreneur. The mutual overlap of these conversations is large by the standard of their overlap with non-economic conversations – you can get any economist to talk to you about the entry of new firms into ecological niches, for example, or the adequacies of a monetary theory of inflation – but the lack of overlap is large, too, by the standard of what it should be.

Speaking of conversations being more or less similar yet having different notions of how to persuade will make a monist angry. A good monist-detection device is to say to him "Truth is plural" and watch the color of his nose. The monist, though, has had his way for too long in the modern world, traveling about from conversation to conversation instructing people in the law. "Intelligence," he says, "must be measured in a single number and be used to stream school children." "The writing of history is solely a matter of gathering pre-existing facts from archives." "Economics must not use questionnaires, because any behaviorist knows that these might be falsely answered." "Economics will only be a real Science when it uses experiments such as a withered branch of psychology once depended on."

The new pluralist and pragmatic and hermeneutic and rhetorical conversation about the conversation "weaves a web of significance," in Clifford Geertz's phrase, around the talk of economists. The new conversation in economics is only imitating what the economists themselves actually do with their stories and metaphors when they talk about the Federal Reserve Board and the trade deficit with Japan. As the great applied economist Sir Alec Cairncross put it,

> When it comes to action, economic theory is only one input among many. It has to be combined with a grasp of political and administrative feasibility and above all has to take advantage of experience and observation, not rely wholly on logic. As has often been remarked, logic can be a way of going wrong with confidence. (Cairncross 1992, p. 20)

Economics, then, might be well advised to step down from the pedestal on which like the woman of the 1950s it fondly imagines it stands. A conversation in modern economics differs from economic journalism but is similar, differs from fiction but is similar, differs from poetry but is similar, differs from mathematics but is similar, differs from philosophy but is similar. There is no hierarchy here, no monist philosopher king reaching into conversations to spoil their tone. I recommend a rhetorically sophisticated culture for economists, in which, as Richard Rorty says, "neither the priests nor the physicists nor the poets nor the Party were thought of as more 'rational,' or more 'scientific' or 'deeper' than one another. No particular portion of culture would be singled out as exemplifying (or signally failing to exemplify) the condition to which the rest aspired." Or as the linguist James D. McCawley puts it, "no particular tradition has a right to speak for humanity as a whole . . . or for 'Reason' as divorced from all the diverse reasoning individuals and traditions of reasoning" (1990, p. 380). The present attitude, at least among those who have not yet felt the doubts of the Frustrated Scientist and the others, is ignorance about the variety of economics and of similar conversations, an ignorance breeding contempt.

Consider as a down-to-earth example the public conversation in the early 1990s about the budget crisis. The budget crisis was and is a real thing, because Gramm–Rudman–Hollings made it so. But as President Bush would have said it was also a word thing. The words make the crisis, too. . . .

In other words, rhetoric is speech with an audience. All speech that intends to persuade is rhetorical, from higher math to lower advertising. In 1991 the Republican rhetoric of the budget crisis intended to persuade an audience of "middle-income" taxpayers, the victims of the bubble in tax rates, sturdy yeomen, it turns out, who were the top 5 percent of incomes. The same wealthy audience was supposed to be persuaded by the Democratic rhetoric, because the audience of the top 5 percent is the politically influential one. The Democratic rhetoric in 1991 and in the election campaign of 1992 was to propose taxing the very (very) rich in order to save the "middle class." "Don't tax him; don't tax me; / Tax that fellow eating brie." It turns out that there aren't enough brie eaters to solve the budget crisis.

But wait a minute. The expert economists offer us a way out of the rhetoric, don't they? The public and politicians indulge in wordcraft, but don't the experts just give us the plain facts and logic?

No, they don't. Experts want to persuade audiences, too, and therefore exercise wordcraft, in no dishonorable sense. Their rhetorics agree on some points. For instance, economists agree that the "crisis" is self-imposed, a weapon wielded by the economist-turned-senator, Phil Gramm, trying to get the mule's attention. But the economic experts disagree on whether the "underlying problem" of the deficit is serious or not. Their disagreements spring not from idiocy or bad faith but from rhetorical choices, often made unconsciously.

Suppose the economist uses a metaphor of the United States as a mere portion of a world economy, in the same way as Iowa is a portion of the upper Midwest. He will therefore not believe the story of the deficit causing a higher interest rate in the United States. The interest rate, he will say, is a result of the whole world's demand for funds. Quit worrying about the little piece of it called the US federal deficit. Or suppose the economist uses a story of a slippery slope to socialism. In that story a loosening of the federal budget leads to B-1 bombers and subsidies to farm owners in the top 5 percent of incomes.

The expertise shows in the rhetoric, though many of the experts don't recognize their own rhetoric. An economist is a poet / But doesn't know it. He is a novelist, too, and lives happily ever after. He is a philosopher, but does not know himself. Is the budget in crisis? It depends on your wordcraft, that Greek word to the wise, "rhetoric."

References

Aristotle. 1991. *Rhetoric.* Trans. George A. Kennedy. New York: Oxford University Press.

Barry, Brian. 1965. *Political Argument.* London: Routledge Kegan Paul.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* National Bureau for Economic Research. New York: Columbia University Press.

Berger, John. 1985. *The Sense of Sight: Writings by John Berger.* Ed. L. Spencer. New York: Pantheon.

Bicchieri, Cristina. 1988. "Should a Scientist Abstain from Metaphor?" pp. 100–114 in Klamer, Solow, and McCloskey, eds., *Consequences.*

Blundell, William E. 1988. *The Art and Craft of Feature Writing.* New York: New American Library.

Cairncross, Alec. 1992. "From Theory to Policy-Making: Economics as a Profession." *Banco Nazionale del Lavoro Quarterly Review* 180 (March): 3–20.

Feiwel, George R. 1987. *Arrow and the Ascent of Modern Economics.* Basingstoke: Macmillan.

Gergen, Kenneth J. 1986. "Correspondence versus Autonomy in the Language of Understanding Human Action." pp. 136–162 in Fiske and Shweder, eds., *Metatheory in Social Science.*

Gordon, David. 1991. "Review of McCloskey's, 'If You're So Smart.'" *Review of Austrian Economics* 5 (2): 123–127.

Kirby, John T. 1990. "The 'Great Triangle' in Early Greek Rhetoric and Poetics." *Rhetorica* 8: 213–228.

Klamer, Arjo. 1983. *Conversations with Economists: New Classical Economists and Opponents Speak out on the Current Controversy in Macroeconomics*, Totawa, N. J.: Rowman and Allanheld.

———. 1984. "Levels of Discourse in New Classical Economics." *History of Political Economy* 16 (Summer): 263–290.

Klamer, Arjo and Thomas C. Leonard. 1993. "So What's a Metaphor?" In Philip Mirowski, ed., *Natural Images in Economics*. Cambridge: Cambridge University Press.

Klamer, Arjo, Robert M. Solow, and D. N. McCloskey, eds., 1988. *The Consequences of Economic Rhetoric*. New York: Cambridge University Press.

Kuhn, Thomas. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Lanham, Richard A. 1974. *Style: An Anti-Textbook*. New Haven: Yale University Press.

———. 1992. "The Extraordinary Convergence: Democracy, Technology, Theory, and the University Curriculum." pp. 33–56 in Darryl J. Gless and Barbara Herrnstein Smith, eds., *The Politics of Liberal Education*. Durham, N.C.: Duke University Press.

———. 1993. *The Electronic World: Democracy, Technology, and the Arts*. Chicago: University of Chicago Press.

Levi, Primo, and Tullio Regge. 1992. *Conversations*. Trans. R. Rosenthal. Harmondsworth: Penguin Books.

McCawley, James D. 1990. "The Dark Side of Reason [Review of Feyerabend's Farewell to Reason.]" *Critical Review* 4 (3, Summer): 377–385.

McCloskey, D. N. 1985. *The Rhetoric of Economics*. Madison: University of Wisconsin Press.

McGrath, Francis C. 1985. "How Metaphor Works: What Boyle's Law and Shakespeare's 73rd Sonnet Have in Common." Department of English, University of Southern Maine, Portland, Maine.

Marcus, Solomon. 1974. "Fifty-two Oppositions between Scientific and Poetic Communication." pp. 83–96 in C. Cherry, ed., *Pragmatic Aspects of Human Communication*. Dordrecht, Holland: Reidel.

Moore, Ruth. 1966 (1985). *Niels Bohr: The Man, His Science, and the World They Changed*. Cambridge, Mass. and London: MIT Press.

Nelson, John. 1983. "Models, Statistics, and Other Tropes of Politics: Or, Whatever Happend to Argument in Political Science?" In Zarefsky, D., Sillars, M. O., and Rhodes, J., eds., *Argument in Transition: Proceedings of the Third Summer Conference on Argumentation*. Annandale, Va.: Speech Communication Association.

Oakeshott, Michael. 1959 (1991). "The Voice of Poetry in the Conversation of Mankind." pp. 488–541 in Oakeshott, *Rationalism in Politics and Other Essays*. Indianapolis: Liberty Classics.

Petrey, Sandy. 1990. *Speech Acts and Literary Theory*. New York and London: Routledge & Kegan Paul.

Plato. 1925. *Georgias*. Trans. W. R. M. Lamb. Cambridge, Mass.: Harvard University Press.

———. 1914. *Phaedrus*. Trans, H. N. Fowler. Cambridge, Mass.: Harvard University Press.

Quine, W. V. 1987. *Quiddities: An Intermittently Philosophical Dictionary*. Cambridge, Mass.: Harvard University Press.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Rosenblatt, Louise M. 1978. *The Reader, the Text, the Poem: The Transactional Theory of the Literary Work*. Carbondale: Southern Illinois University Press.
Thucydides. 1972. *History of the Peloponnesian War*. Trans. R. Warner. Harmondsworth: Penguin Books.
Thurow, Lester. 1985. *The Zero-Sum Solution: Building a World-Class American Economy*. New York: Simon and Schuster.
Wittgenstein, Ludwig. 1945 (1958). *Philosophical Investigations: The English Text of the Third Edition*. Trans. G. E. M. Anscombe. New York: Macmillan.

TWENTY-THREE

Realism

Uskali Mäki

Uskali Mäki (1951–) is currently an Academy Professor in the Academy of Finland. He holds a Ph.D. from Faculty of the Social Sciences at the University of Helsinki and has published extensively in both economics and philosophy. The author of over one hundred essays and a past editor of *The Journal of Economic Methodology*, Mäki's interests span the entire domain of economic methodology. Realism has been a persistent interest, and a number of his publications explore the varieties of realism to be found in economics. Mäki has played a key role in establishing economic methodology as a discipline.

When an economist talks about the 'realism of assumptions', he is not using the term 'realism' in any of its standard philosophical senses. Another difficulty that plagues the term is that it has a variety of legitimate philosophical meanings that are interrelated but do not reduce to each other. 'Realism' is used as the name for a variety of doctrines about things such as science, sense perception, universals, other minds, the past, mathematical objects, truth, moral values, possibilities and so on. This is expressed in the fact that the opponents of realists on these issues are not called uniformly by a single label. Depending on the issue at hand, the non-realists are said to subscribe to positions such as idealism, phenomenalism, empiricism, nominalism, conventionalism, instrumentalism, operationism, fictionalism, relativism and constructivism. This variety is also the reason why no shorthand definition – and no single non-disjunctive definition, whether short or long – of the term 'realism' can be provided. The following considers ontological, semantic and epistemological formulations of realism without pretending to be exhaustive.

Pages 404–09 of *The Handbook of Economic Methodology*, edited by John B. Davis, D. Wade Hands, and Uskali Mäki. Cheltenham, UK: Edward Elgar.

As an *ontological* doctrine, realism has the general form of the statement, '*X* exists', or '*X*s are real'. '*X*' is a variable which may acquire different specifications. For each specification there corresponds a version of ontological realism. The most general and weakest variety of ontological realism is produced when '*X*' is replaced with 'the world'. No further specifications of the constituents and nature of the world is provided. This form of ontological realism does not include any ideas about the way the world exists, it only amounts to the idea *that* the world exists.

If we replace '*X*' with 'universals', we get doctrines such as *Platonic* or *Aristotelian realism* which state that universals exist; that is, it is universals (alone or also) that constitute the world. Not only (or not) the many particulars, such as round objects and business firms and rational men, but also (or only) roundness and firmhood and rationality and manhood exist. This is in fact the original usage of 'realism', used in connection with the debate over universals between the realists and their opponents, the nominalists. Nominalism may also be a form of realism: it replaces '*X*' with 'particulars' and states that there is nothing but particulars in the world.

If we replace '*X*' with 'medium-sized material entities' or 'objects of sense perception', we get standard forms of ontological *commonsense realism*. In general, realist theories in this category state that the perceivable commonsense world is real; in other words, that objects such as clouds and clocks, horses and houses, mountains and marmalade exist – that the objects that common sense takes to exist do exist in the objective way that common sense takes them to exist. The opponents of realism in the theory of perception include the idealists (to whom it is 'ideas' that constitute the world) and phenomenalists (who try to construe the world out of what they call 'sense data'). Commonsense realism can also be taken to comprise the idea that the mental entities in terms of which 'commonsense psychology' or 'folk psychology' conceptualizes our lives and behaviour exist. Accordingly, there is a fact of the matter regarding what we intend, want, believe, mean, hope and fear; that is, intentions, wants, beliefs and meanings exist (even in cases where what they appear to be about do not exist – such as mermaids and Rudolph the Reindeer). Eliminative materialists are among the opponents of commonsense realism about the mental.

If we replace '*X*' with 'the (often unobservable) entities as objects of (most or best, current or future) scientific theories' such as electrons, photons, quarks, electromagnetic fields, curved space–time, genes, viruses, brain states and so on, we get the ontological statement of *scientific realism*. The world as postulated in scientific theories now becomes the (or a) world that is real. The opponents of realism about the ontology of scientific theories are the fictionalists and (ontological) instrumentalists. In the case of radical physicalist scientific realism – according to which only entities postulated by physical sciences exist – the opponents also include those who advocate commonsense realism. More moderate forms of scientific realism may accommodate the existence of perceivable material entities of the commonsense world or even (at least some of) the mental entities of commonsense psychology.

There are other and more controversial versions, such as the one we get by replacing '*X*' with 'possible worlds', called *modal realism* by its advocates, such as David Lewis. According to this version, existence is not restricted to the actual world; the actual world is just one among many existing possible worlds.

An important question concerns the *quantifier* that the above forms of realism could use in relation to the entities that are claimed to exist. Such a quantifier indicates answers to the question 'how many?' such as 'all', 'no' and 'some'. No realist would like to claim that *all* posited universals, particulars, commonsense objects and/or scientific objects exist (this would imply that Father Christmas, centaurs and phlogiston are all as real as green tealeaves and DNA molecules). Many other realists commit themselves to the existence of at least some of these entities. Many scientific realists would say that *most* of the objects postulated in well-established scientific theories exist. However, none of these quantifiers is necessary for ontological realism; none of them should be included in the definition of realism. It is sufficient for realism about *X* to hold that *X* might exist, that the notion of *X* existing is a sensible and coherent notion. This raises the key issue concerning the appropriate concepts of existence.

Specification of the types and numbers of entities that are claimed to exist is not sufficient for a complete understanding of ontological realism. It is necessary to specify what is meant by the expressions 'exists' and 'is real'. The first decision to be made about this question concerns whether there is an ontological notion of 'exists' and 'is real' that is conceptually independent of epistemic considerations and of specific conceptual frameworks in which the specifications of 'X' appear; that is, whether in addition to a concept of 'existence within a framework' there is a sensible and 'X framework-independent' concept for expressing the idea that X exists (or does not exist). For example, the question may be whether the meaning of 'exist' in the statement, 'photons exist', may be understood independently of the meaning of the term 'photon' and of the specific theoretical frameworks in which 'photon' is embedded and of the epistemic claims we may feel to be justified in making about the existence of photons (such as 'the evidence suggests that photons exist'). If we claim that the very ontological notion of existence is in such a way independent, our notion is unproblematically a realist one. There are those, however, who claim to be realists but deny the framework-independent notion of 'exists' and 'is real'; they sometimes call themselves 'internal' or 'epistemic' realists, while 'external' or 'metaphysical' realism is preserved for those who subscribe to the independence thesis.

The second decision concerns the meaning of 'exists' and 'is real' more directly. The conventional specification is '*exists mind-independently*' or '*exists independently of the human mind*'. This formulation has the implication that it excludes realism about mental entities and entities dependent on the mental, such as persons, material artefacts, and social institutions construed in a dualist or non-eliminative physicalist fashion; it restricts the scope of realism to the material or physical world. What this entails is materialist or physicalist realism. Obviously, this formulation is not able to accommodate realism about social sciences postulating things such as intentions, expectations, roles, conventions or institutions; thus economics would be a hopeless case for such a realist. There are alternative specifications of 'exists' that could be thought to avoid the above implication, such as '*exists recognition-independently*' or '*exists inquiry-independently*' or '*exists independently of any particular act of representation of it*'. One may then argue that mental entities and/or social entities exist in one or more such senses and that these senses are genuinely realist ones.

'Realism' has increasingly also become a name for some *semantic* views, that is, views concerning such things as reference and truth. In the formulations above, the use of the notion of reference could not be avoided completely; for example, scientific realism as an ontological thesis about the existence of certain objects was defined in terms of theories being about those objects, that is, theories referring to them. The claim that scientific theories and the terms they include *refer to* real existents is part of the semantic thesis of scientific realism. The other part of semantic realism is the thesis that the sentences contained in scientific theories are genuine, *true or false*, statements about the real world and that they have a truth value irrespective of whether we are able to determine it. Some philosophers, such as Dummett, take bivalence – the principle that every proposition is either true or false – as a defining characteristic of realism. Standard forms of instrumentalism are among the opponents of realism so understood. For them, scientific theories are just calculation or inference devices with no semantic ties to reality; or if they have semantic properties – such as that of falsehood – these properties are taken to be irrelevant for our assessment of them. Various epistemic conceptions of truth, such as the idea of truth as warranted assertability or truth as idealized rational acceptability, challenge the notion that truth may escape even ideal knowers in ideal conditions; that truth is independent of our ways and chances of finding out about it. Pragmatisms of various sorts contain the negation of this realist idea. As against such views, realists hold the view that even a methodologically perfect theory, fully satisfying all the desiderata we can imagine, can still be mistaken.

These ideas are sometimes complemented by other views attributed to realism about science, such as that most current scientific theories are (at least approximately) true and/or that, as science develops, its theories get progressively nearer to the truth. However, these ideas are not required by the most basic and simple realist theses; realism might be correct even though most current science was wrong and even if science did not converge towards the truth. The same could be argued to be the case with a popular normative idea ascribed to realism about science, namely that science should pursue true accounts of the world.

'Realism' is also used in connection with a specific view of what truth is, namely the correspondence theory of truth. There are many versions of this theory, but they share the idea that it is somehow partly in virtue of the way the world is that sentences (statements, beliefs or utterances) are true or false – that truth is in this sense objective – and that truth amounts to a correspondence between the way the world is claimed to be and the way the world is. Correspondence theories differ from one another as to how they view the correspondence relation and the two poles linked by it. Many realists wish to link this view of truth with the very concept of realism. Many others dispute the connection. Some of them endorse the correspondence theory as a separate idea which is not implied by realism. Others again try to do without the correspondence theory, typically favouring redundancy or deflationary or minimalist views of truth instead; these views imply that, roughly speaking, to say that 'it is true that free trade tends to equalize factory prices' adds nothing to 'free trade tends to equalize factor prices'.

While the ontological meaning of 'realism' is the traditional and primary one, it can be and has been complemented by the epistemological point that the *X*s that are claimed to exist are also knowable. Different forms of *epistemological realism* presuppose some versions of ontological realism and semantic realism and add to them the idea of being known or being knowable. Epistemological realism says of some existing *X* that facts about *X* are known or can be known, implying that knowers have epistemic access to *X*, that there is no veil separating the cognitive subject and the existing object. A variety of different forms of epistemological realism is possible, depending on how one analyses the very idea of knowledge, how the means of knowledge acquisition are conceived and what the relevant *X* is taken to be. Regarding the latter two matters, one can be a scientific epistemological realist, relying on what science now claims to be the case or on the potential capability of science to find out facts about the world, where both cases are based on a reliance on the cognitive power of the theoretical and empirical procedures (as well as the institutional organization) used by science.

Likewise, one can be a commonsense realist about perceptual knowledge. Traditionally, theories of perception have played a dominant role in realist epistemologies. Realist doctrines about perception are usually divided into two main categories, direct and indirect realism. *Direct realism* says that perception is directly about (is a direct awareness of) material objects which exist and that nothing else exists between perception and perceptible objects. *Naive realism*, usually unsupported by philosophers but postulated for purposes of criticism, is a version of direct realism which states that we perceive objects as they are; that is, that sense data or sensible qualities are the intrinsic properties of material objects and that these objects have all the properties they are perceived as having and that these properties are not affected by changes in perceivers and conditions of perception. *Indirect realism* states that perception is directly about mental representations (such as bodily sensations and after-images) and only indirectly about the external world and that both its direct and indirect objects exist. Versions of indirect realism include Locke's *representative realism* and a movement in the United States in the first decades of the twentieth century known as *critical realism* (Lovejoy, Santayana, R. W. Sellars and others). In a generalized form, the label 'critical realism' has been adopted by philosophers to indicate the view that there is a difference between that which is experienced and that which exists independently of being experienced. More generally, philosophers advocating such views emphasize the possibly distorting contribution of the knowing subject to the cognition.

This is related to the idea of fallibility. Fallibilism is the view that knowledge claims are in principle fallible (and possibly corrigible, revisable in the light of further evidence and arguments) so that full certitude is unattainable. Realists typically are fallibilists, opposing both dogmatism and radical scepticism. Even more, realism is often defined so as to presuppose fallibilism. This is entailed by the idea, mentioned above, that even an epistemically ideal theory or statement may be wrong.

It is a widely shared view among realist philosophers that the resolution of (or to put it less strongly: progress with respect to) issues about many themes mentioned above, such as what the world is made of, and what reference, truth and knowledge amount to, is up to future science. In other words, the specifications are understood as being *a posteriori* in regard to the progress of special sciences such as biology, cognitive science and, to anticipate boldly, economics. It is not the task of philosophy, in this opinion, to decide *a priori* what kinds of entities exist, what structure the world has, what relations our language has to the non-linguistic reality, what can be known and perceived, and so on. As an *a posteriori* exercise, philosophy produces claims that are fallible in the same sense that any other claims may be wrong.

Do realism and economics fit together? This is a question of interest to economists, economic methodologists, philosophers of science, politicians and lay public. The answer depends on what we mean by 'economics' and 'realism'. For example, we can take 'economics' to refer variously to any current form of economics or to economics as we would like it to be or economics as it might be – and the answer might vary accordingly. As for realism, we have not exhausted the full list in the foregoing, but we have come up with many forms of realism, and the answer obviously depends on the form(s) we choose. For example, if we opt for radical physicalist scientific realism, current economics will not fit. The outcome is the same if realism is taken to require that all components of economic theories be true.

However, a number of economists have been shown or can be shown to subscribe to one or another form of realism. These include J. H. von Thünen, J. S. Mill, Karl Marx, J. E. Cairnes, Carl Menger, Lionel Robbins, Nicholas Kaldor, Milton Friedman, Ronald Coase, George Richardson, Oliver Williamson and others. Even though there are important differences between them, they share the view that economic reality has an objectively (albeit not mind-independently) existing structure, and that economic theories, even though being partial and involving false elements, are able to truly represent some of the important aspects of this reality. There are some special features regarding realism about economics, such as commonsense realism playing a prominent role. This is because economic theories much of the time appear to be pretty much about the same objects that our commonsense understanding of the economy is about, such as households and business firms, money and prices, buying and selling, wants and expectations. Another feature, and an epistemologically significant one, is that the simplified and isolated settings theoretically brought about by economists usually cannot be reproduced empirically, thus making the empirical testing of truth claims particularly difficult. Fallibilism should therefore play an exceptionally prominent role in economics.

As an explicit research project, realism has been explored only recently in economic methodology. Two major realist projects have been those of Uskali Mäki (the first published statement appeared in 1982) and Tony Lawson (for example 1997). The differences between these two projects are many, but two 'meta-methodological' ones stand out immediately. One is that Lawson's project is largely an application of one philosophical system, that of Roy Bhaskar, while Mäki's is a matter of drawing from different philosophical sources as well as creating new conceptual tools that are hoped to reflect some of the peculiar features of economics. The other is that Lawson's project is supposed to have more or less direct critical implications about the poverty of what is called mainstream economics, while Mäki's project has been more neutral: the normative implications are expected to be more indirect and to require lots of factual premises that go beyond realism as a philosophical doctrine. Boylan and O'Gorman (1995) provide expositions and criticisms of these two projects. These projects do not exhaust all there is to the study of realism in the context of economics. Many other economic methodologists and philosophers of economics (such as Alex Rosenberg, Alan Nelson, Daniel Hausman, Don Ross, Nancy Cartwright and others) have contributed to the realist project without necessarily doing it explicitly under the banner of 'realism'.

References

Armstrong, David (1978), *Universals and Scientific Realism*, vols I–II, Cambridge.

Boylan, Tom and Paschal O'Gorman (1995), *Beyond Rhetoric and Realism*, London: Routledge.

Devitt, Michael (1984), *Realism and Truth*, Princeton: Princeton University Press.

Lawson, Tony (1997), *Economics and Reality*, Cambridge: Cambridge University Press.

Leplin, Jarrett (ed.) (1984), *Scientific Realism*, Berkeley: University of California Press.

Mäki, Uskali (1989), 'On the problem of realism in economics', *Ricerche Economiche*, **43**, 176–98; reprinted in Bruce Caldwell (ed.), *The Philosophy and Methodology of Economics*, Aldershot: Edward Elgar, 1993.

Mäki, Uskali (1992), 'On the method of isolation in economics', *Idealization IV: Intelligibility in Science*, ed. Craig Dilworth, special issue of *Poznan Studies in the Philosophy of the Sciences and the Humanities*, **26**, 319–54.

Mäki, Uskali (1993), 'The market as an isolated causal process: A metaphysical ground for realism', in Bruce Caldwell and Stephan Boehm (eds), *Austrian Economics: Tensions and New Directions*, Dordrecht: Kluwer.

Mäki, Uskali (1996), 'Scientific realism and some peculiarities of economics', *Boston Studies in the Philosophy of Science*, **169**, 425–45.

Putnam, Hilary (1981), *Reason, Truth and History*, Cambridge: Cambridge University Press.

Searle, John (1995), *The Construction of Social Reality*, New York: Free Press.

Sellars, Wilfrid (1963), *Science, Perception and Reality*, London: Routledge & Kegan Paul.

TWENTY-FOUR

What Has Realism Got to Do with It?

Tony Lawson

Tony Lawson (1950–) received a Ph.D. in Economics from Cambridge University after a first degree in mathematics from the University of London. Lawson is the organizer of the long-running Cambridge Realist Workshop and the Cambridge Social Ontology Group, and he is currently executive director of the Cambridge Centre for Gender Studies. His work lies on the boundaries between philosophy and economics, with a special emphasis on ontology. He has been an editor of the *Cambridge Journal of Economics* for twenty-five years, and sits on many other editorial boards including that of *Feminist Economics*. The author of numerous papers and books, Lawson has played a key role in promoting heterodox economics and establishing social ontology as a focus of research in modern economics.

For several years now I and a number of others (see e.g., Fleetwood, 1999) have been contributing to a project in economics that is often referred to as realist. In a recent article in *Economics and Philosophy* Dan Hausman questions whether realism is actually a feature of this project worth emphasizing (Hausman, 1998). In fact, Hausman goes as far as to suggest that making reference to this aspect of the project may actually be misleading or otherwise unhelpful. His basic worry is summarized in the concluding section of his paper where he writes:

> To label one's program for economic methodology as 'realist' inevitably suggests that the competing programs are not realist or fail to be realist enough. In the case of economic methodology, this suggestion is misleading, because there is no anti-realist school of economic methodology, and there are few methodologists (as opposed to economists) who are instrumentalists either. What is distinctive about Lawson's and Mäki's programs is not realism – which they share with the rest of

For helpful comments on an earlier (longer) draft I am grateful to Clive Lawson, Paul Lewis, Stephen Pratten and Jochen Runde.

Economics and Philosophy, vol. 15 (1999): 269–82. Reprinted with the permission of Cambridge University Press.

economic methodology – but something else. That something else can, of course, be a particular formulation of realism, such as Lawson's critical realism. But it would be less misleading if what was distinctive was characterized in terms of what distinguishes it from alternatives, rather than in terms of what it shares with them. (pp. 208–9)

Now I infer from this passage that Hausman's concern lies not with the project in question being *interpreted* as realist (for Hausman acknowledges that it is); nor necessarily (or primarily) with its being *distinguished* or *identified* as a *specific formulation* of realism (after all Hausman accepts that the 'something else' that can make a programme 'distinctive' 'can, of course, be a particular formulation of realism, such as Lawson's critical realism'); but with it sometimes being *distinguished* or *identified* simply as *realist*. For it is this practice before any other that '*inevitably* suggests that the competing programs are not realist or fail to be realist enough' (emphasis added). In truth, however, I believe that even this latter practice is more than justified in the circumstances. My primary object here is briefly to indicate why. In the course of pursuing it I take the opportunity to identify what I believe to be some significant differences between Hausman's programme and my own.

Realist as a Contrast to Non-realist

Notice first that Hausman identifies two possible inferences to be drawn where a project is identified explicitly as 'realist': that competing programs either 'are not realist', or 'fail to be realist enough'. Let me consider each in turn.

Now I am, of course, ready to acknowledge that we are indeed all realists of sorts, and that this is so with regard to many interpretations of the term. Indeed I have usually taken pains to emphasize as much. In *Economics and Reality*, for example, I acknowledge that '. . . any position might be designated a realism (in the philosophical sense of the term) that asserts the existence of some disputed kind of entity . . . ' (Lawson, 1997a, p. 15). And I add that, 'Clearly on this definition we are all realists of a kind, and there are very many conceivable realisms' (p. 15). I believe, too, that most scientists are scientific realists regarding (are committed to the independent or prior existence of) at least some of their posited objects of investigation.

To be sure, I usually use the term realism in a specific way, primarily to indicate an ontological orientation. However, I am equally ready to acknowledge that we are all realists even in the particular sense of holding (or at least presupposing) ontological positions. In my writings, I have continually acknowledged that all methods and criteria, etc., presuppose an implicit ontology, an unstated account of reality (see for example Lawson, 1997a, p. 49). Even explicit attempts to suppress ontology result only in the generation of an implicit one, as I have frequently pointed out in discussing Hume's empiricism or forms of postmodernism (see, for example, Lawson, 1997a, Chapter 6).

So we are all realists of some sort, and perhaps of many sorts. Does it follow thereby that a social-theory project oriented to economics should not identify itself as realist? Not at all. Specific projects, programmes and activities in all walks of life are regularly identified according to certain fundamental aspects or features which also figure in other projects, etc., but less centrally so.

For example, Hausman distinguishes his own project as one in methodology. I, Mäki, and others whom Hausman identifies explicitly, are all – and are interpreted by Hausman as being – involved with methodology in our projects. And so indeed are all economists. All approaches, methods, techniques, goals, criteria, etc., adopted by economists, and everyone else, presuppose conceptions of scientific or proper method. Thus all economists, and indeed all scientists or researchers, are inherently methodological. Does this mean that Hausman should desist from designating his particular project as explicitly methodological, in case he is erroneously interpreted as implying thereby that all other projects are not methodological? Are cooks not to be distinguished as cooks because we all cook? Or singers as singers, runners as runners, economists as economists, teachers as teachers, students as students, and so on?

Realist: More Rather Than Less

The reason most of us do not distinguish ourselves as cooks, singers, runners, carers or cleaners, etc., is not because we do not engage in these activities, but because we do not pursue any of them sufficiently ardently or seriously, on a regular or consistent enough basis. This, I think, is the relevant point. And it brings me to Hausman's second worry about explicitly labelling a project realist: that so doing 'inevitably suggests that the competing programs . . . fail to be realist enough'. There is a sense in which this is exactly what I *am* suggesting.

In identifying my project as realist I am first and foremost wanting to indicate a *conscious* and *sustained* orientation towards examining, and formulating *explicit* positions concerning, the nature and structure of social reality, as well as investigating the nature and grounds of ontological (and other) presuppositions of prominent or otherwise significant or interesting contributions. And I am wanting to suggest that it is precisely this sort of *explicit concern* with questions of ontology that is (or has been) lacking in modern economics. This is an absence, indeed, that I believe contributes significantly to the discipline's current malaise. In this sense of the term, in my view, most of the projects contributing to the development of modern economics are not nearly realist enough.

But there is, at least, a second, albeit related, sense in which projects in economics are not realist enough, and in which I want to distinguish my own. I refer here to the tendency of most projects in economics implicitly to conflate the real with what is or could be apparent (or is 'observable'), and fail in any meaningful, systematic or sustained fashion to go beyond appearances to the (equally real – but not necessarily wholly apparent) underlying structures, powers, mechanisms and tendencies that generate or condition the surface phenomena of reality.

Competing Programmes

Before giving more detail on all this, however, I must refer once more to Hausman's specific formulation of his worry: that others may infer that I am suggesting 'that the competing programs are not realist or fail to be realist enough'. The feature of it I want to focus upon here is Hausman's reference to 'competing programs'. It seems from the longer passage in which this inference is embedded (noted at the outset above), that Hausman has in mind here only programmes of methodologists, where methodologists are explicitly contrasted, by Hausman, to economists. I should immediately emphasize, then, that Hausman is in error if he is presuming, as he seems to be, that, in adopting such a realist orientation, the 'competing programs' I am addressing are only, or even primarily, those of others who explicitly distinguish themselves as methodologists. I do address these. But they are by no means my main intended audiences, opponents or 'competitors'.

Economics and Reality, for example, is explicitly aimed at an audience of general economists, both mainstream and non-mainstream. I view it as a contribution to economics as social theory. The identified 'opponents' are contributors, or potential contributors, to the contemporary mainstream programme (p. xvii). I optimistically saw (and continue to see) myself as attempting to contribute to a fruitful transformation of the discipline. And

I saw, and see, the taking of an *explicitly* realist orientation as a significant step to this end (p. 15).

That said, it happens that I do believe that most contributions explicitly designated programmes in economic methodology are (or have been) also not realist enough. I return to this below. For the time being, however, I want to concentrate on the mainstream programme in economics. For, as I say, I believe that this is indeed not realist enough in the senses indicated, and that the contrast between this project and my own is alone sufficient justification – although not the only reason – for distinguishing the latter explicitly as realist.

The Problem with Modern Mainstream Economics

Let me be more specific. First, for anyone with an interest in metaphysics, it does not take too much familiarity with our discipline to recognize a continuing failure of modern economists to examine the nature and consistency, etc., of the ontological, and other, presuppositions of their various pronouncements on, and decisions concerning, matters of substantive theory and method. It follows immediately, I think, that any project concerned systematically so to examine these ontological questions can fairly be *identified* explicitly as a realist. My own project can certainly be identified as realist in this sense.

But I do go much further in a direction that can (also) be reasonably described as realist. Specifically, I, along with others, have been engaged in a range of investigations aimed at attempting to provide a sustainable social metaphysics, a theory of social reality, to inform the fashioning of methods of social, including economic, analysis. Fundamental to this project has been the questioning of the nature of social material, investigating its mode of being, structure and peculiarities, conditions, and so forth.

The orientation of this project thus contrasts quite significantly with the mainstream approach of (unthinkingly) adopting methods assumed to be successfully utilized in the natural sciences or somehow thought, on an a priori basis, to characterize proper science. Fundamental to the mainstream position is an insistence on working with formalistic models. Indeed, the primary objective of this mainstream tradition is the production of theories that facilitate mathematical tractability. In contrast my goal, naive though it may sound, is to pursue true theories, or at least to achieve those that are explanatorily powerful.

I have found that the two sets of objectives, explanatorily powerful theories and mathematically tractable models, are usually incompatible, just because of the nature of the social world. For whereas the latter has been found to be quintessentially open and seemingly susceptible to scientifically interesting local closures, the generalized use of formalistic economic methods presupposes that the social world is everywhere closed. By a closure here I mean merely a system in which event regularities (deterministic or probabilistic) occur.

Two sets of observations, at least, can be explained by this incompatibility: that formalistic models are rarely, if ever, found to be empirically successful, and that the entities posited in mainstream theorizing (determined on the basis of being facilitative of mathematical tractability and other mainly pragmatic criteria) are usually seen to be unrealistic in many essential ways.

Mainstream economists seem to suppose there is no other way of proceeding. For those econometricians – most if not all of them – who care about 'fit' with reality (assumed to be captured by data on measurable events and states of affairs) the hope appears to be that empirically successful models will be uncovered in due course. But although some equally hope the positing of theoretical entities recognized as severely unrealistic is also a temporary situation, many act as if it does not matter. Here there are indeed *elements* of anti-realism in the mainstream position. Some are quite *explicitly* anti-realist when commenting on the perceived status of the 'theoretical entities' that are used to ground their models (if not in their conceptions of the events and states of affairs thought to be captured by the 'data'). That is, they explicitly pronounce, a philosophical theory – about economic theories or models – to the effect that either it is not meaningful to talk of true models or true models are not possible (for example, see the discussion in Lawson 1997a, especially page 325).

Now the primary problem or error of the mainstream project here, as I see it, is not the anti-realist orientation of many of its participants towards formalistic economic models *per se* (a stance which would appear to carry some justification), but the decision to persevere with – and to insist that all economists concern themselves with almost nothing but – the modelling project despite its long-term and continuing lack of clear empirical successes. The central mistake is one of not recognizing that the near exclusive focus upon closed systems modelling – a procedure mainly suited to certain natural (experimental) contexts – is itself questionable, and in need of justification. This, I believe, is the key to the mainstream discipline's shortfalls, turning on the more general avoidance of an explicit concern with ontology, of omitting to investigate the nature of social reality with a view to determining the basis of potentially more fruitful alternatives. In a word, the primary failing of modern economists is ontological neglect. It is in this specific sense especially, in my view, that most economists are not being realist enough.

A Realist Alternative

Of course, the sorts of responses by modellers just noted mainly serve the purpose of allowing the modelling project in economics to continue unabated. In some quarters numerous pragmatic or coherentist criteria of model selection (elegance, parsimony, complexity, consistency with the equilibrium framework) are even invoked as if to obviate any need for the empirical assessment of models.

Such responses are certainly questionable. My alternative strategy, as I say, has been to investigate in a sustained and explicit way the nature of social reality and to tailor methods of social investigation accordingly. This has certainly led me to doubt whether methods of formalistic analysis have much relevance to the social domain. And in the process I have come to defend a conception of the social realm as emergent from, but irreducible to, human interaction. I have argued for a theory of social ontology that includes forms of social structure, including social relationships, rules, positions, processes and totalities, etc., that collectively constitute a relatively autonomous realm, being dependent upon and resulting from human interaction, but with properties that are irreducible to human interaction, though acting back upon it. In *Economics and Reality* I argue, in effect, that this social ontology covers both a 'vertical realism', entailing a commitment to underlying social structures, powers and entities, etc., and also a 'horizontal realism' covering the transfactual operation of causal mechanisms in open and (any conceivable) closed systems alike, that is, whatever the outcomes. In this I find that causally efficacious (and often largely unobservable) social structures and mechanisms, etc., indeed exist independently of our investigations of them and, individually and collectively, constitute proper objects of social scientific study.

The Situation in 'Economic Methodology'

To this point I have been referring mainly to the ontological neglect (in the sense of failing to sustain explicit ontological reasoning as well as ontological depth) and the consequences of this for the methodological practices of mainstream economists. In comparison to the largely a priori unthinkingly reductionist and scientistic programme perpetrated by mainstream economists, then, the project in social theory to which I have contributed can, I believe, with good reason be *identified* as realist.

Hausman, though, appears to be more concerned with any implied contrasts with other projects explicitly designated methodological. Now if it is the case that such methodological projects in economics are concerned to engage in a significant way in social metaphysics explicitly, these projects, in my view, warrant *being identified explicitly as realist as well.* Certainly, I take this to be appropriate where any such project is concerned *in this manner* to confront and ultimately help transform (and so inevitably be contrasted with) the largely a priori scientistic set of practices that is the contemporary mainstream. Moreover, if in the course of such a critical endeavour a definite perspective on the nature of nature, science and society were derived and defended, any such project would equally warrant being identified as a *specific formulation* of realism.

As it happens, however, I do believe that those projects in modern economics explicitly designated methodological have mostly (with a few exceptions) also failed to give sufficient attention to questions of ontology or metaphysics. In the main it is questions of epistemic appraisal (i.e., epistemological questions concerning the rational basis for accepting or rejecting theories) that have occupied the economic methodology discussion (for discussions see Lawson, 1997a, pp. xiii–xvi; Fleetwood, 1998, pp. 127–35).

I cannot survey the contributions of these projects here. However, in order to address some further rather important issues raised by Hausman, and to take the discussion further in the hope of bringing clarity to our differences, I might add at this point that I harbour doubts that even Hausman's own programme – highly productive and insightful though it is – pays sufficient attention to questions of metaphysics. It is more oriented to issues of metaphysics than most, and seems to be becoming increasingly so. Yet it seems to me that it, too, ultimately fails – so far – both to challenge sufficiently the relevance (including ontological presuppositions) of the contemporary mainstream – or any other – programme, or to elaborate an ontology that takes us very far beyond the course of actual events and states of affairs. I believe these claims can be shown to be true of Hausman's output broadly conceived. But for present purposes, let me concentrate on the *Economics and Philosophy* paper in question where Hausman, by explicit intent, is actually wanting to address issues of concern to realists.

Hausman and Economics

At various stages in his *Economics and Philosophy* paper, in the course of establishing some point or other (e.g., that 'the debate between realists and epistemological anti-realists is largely irrelevant to economics' (1998, p. 185)), Hausman makes assertions to the effect that 'economics does not postulate unobservables in the way physics does' (1998, p. 185). Indeed, a central thesis is formulated as follows:

> . . . the ontological, semantic, and epistemological issues separating realists from anti-realists and from some instrumentalists, are largely irrelevant to economics. The reason is simple: economic theories for the most part do not postulate new unobservable entities. (1998, p. 196)

The question I want to pursue, of course, is which economic theories are we talking about? All possible? Those formulated in heterodox approaches? Economists who have contributed to, or who have been informed by, the project of critical realism in economics have, in their more substantive contributions, generated economic theories that posit a variety of novel entities which in large part, at least, are unobservable. These include particular social relations (gender, race, employer/employee, student/teacher, money . . .), other structures of power, social processes, social positions, social rules, evolving totalities, specific institutions, etc. (see, eg., Lawson, 1997b, Chapter 18). This research, like that of others on industrial districts, regions, collective learning and so on, is constantly positing new categories, relations, processes and totalities, etc., many, albeit not all, of which are, or possess essential aspects that are, inevitably unobservable. Indeed, human society itself can only be known, and not seen, to exist. Hausman is thus quite wrong when he supposes that I 'would not . . . dispute the claim that economic theories rarely posit the existence of new unobservable entities' (1998, p. 202). (For my conception of the nature of economics specifically see, especially, Lawson 1997b.)

Why should Hausman conclude, despite everything, that 'economic theories for the most part do not posit new unobservable entities'? It may be because, at least at the relevant stage of his discussion, he is implicitly and unquestioningly reducing economic theory to the output of the current mainstream project in economics (or, even worse, to a specific strand of it, to something like mainstream 'theoretical' microeconomics). This, of course, is Hausman's explicit strategy in his recent (1992) book *The Inexact and Separate Science of Economics.* And it is in this light that we can most easily interpret his almost exclusive focus in the *Economics and Philosophy* paper on the question of whether 'the preferences and expectations that explain and predict choices are unobservable' (1998, p. 196). For only to the contemporary mainstream do such matters assume such a central, almost exclusive, role.

Of course, even if we focus only on this limited domain we are entitled to ask why, or in what sense, it matters that unobservables are or are not *new* (apparently meaning unfamiliar or non-commonsensical). For at one point Hausman accepts that beliefs and preferences, etc., are indeed *un*observable and even contested, but seems to suggest that any realist/anti-realist debate this might facilitate in economics is somehow less significant than the debates of physics just because the noted unobservable items are known to us:

> The point I want to insist on is a different one. Anti-realists seek to draw a line between the relatively unproblematic claims of everyday life and the problematic theoretical posits of science. Physics postulates new unobservables, to whose existence commonsense realism does not commit us. Although economics refers to unobservables, it does not, in contrast to physics, postulate new ones. Its unobservables – beliefs, preferences and the like – are venerable. They have been part of a commonsense understanding of the world for millennia. (1998, pp. 197–8).

And he adds below:

> There is no issue concerning realism versus anti-realism in economics that is not simultaneously an issue concerning the everyday understanding of the world. (1998, p. 198)

Now I concur with the latter remark. But I draw from it more or less the opposite inference to Hausman. Certainly, I do not take it to entail, as Hausman mostly does, that we should refrain from questioning the reality and nature of certain aspects of the social realm, just because there exists a commonsense understanding of them (that the unobservables are not in this sense 'new ones'); I do not suppose that 'the everyday understanding of the world' is incorrigible. Rather I believe we should be continually reassessing even the most familiar of our everyday categories. For example, I take money (a feature of everyday modern life with presumably a commonsense understanding) to be a system of social relations (explaining why this piece of metal/paper/plastic functions differently from others). Is this interpretation part of commonsense understanding? Is it to be discounted if (and just because) it is not?

What, too, of the everyday gender-differentiated or class-differentiated, etc., practices, rights, obligations, etc. in any given location and their structural conditions? I even take Hausman's tables and chairs to be constituted in part by social relations. When I go camping, for example, numerous items have the potential to serve as tables (flat topped tree stumps, smooth-sided lumps of rock) or serve as chairs (rocks, upturned buckets, etc.). Which items become so constituted depends in part on us and our relations to them. Ultimately, of course, this is no less true of the artifacts that we call tables and chairs in the home.

In other words, I am suggesting that just as Hausman allows that, 'Physics postulates new unobservables, to whose existence commonsense does not commit us' so too can, and often does, social science including economics. Now this perspective cannot be ruled out just because it allows that we can, and often ought, to transcend or transform commonsense. And in that it interprets or recognizes daily life as internally-related to underlying structures (including wider totalities), I think the perspective sustained does warrant being interpreted as rather more realist than Hausman's somewhat commonsensical and quasi-actualistic (reducing reality to the actual course of events) account.

Still these sorts of considerations are not my only or even my primary concern here. I return to my main point that by mostly focusing on items such as preferences and expectations, Hausman appears implicitly to be interpreting economics as little more than the current, and hardly illuminating, mainstream set of contributions. It is true that Hausman includes in his paper a sub-section with the more promising heading: '*Other unobservables in economics*'. But in essence only two sets of items – 'socially necessary labour time' on the one hand, and 'human capital' and 'attributes' on the other – are identified as real possibilities. And each is quickly dismissed merely on the supposed grounds of being either not 'economically significant' (p. 200) or 'relatively unimportant in economics' (p. 202). But we are once more entitled to ask what is this economics in which these (and other unobservables) are unimportant? It can only be the contemporary mainstream. This is precisely the project I have criticized for its ontological neglect, and I worry that Hausman may be unwittingly colluding in this by taking the output of that largely bankrupt project as sacrosanct.

Now it may seem that against this latter criticism at least, Hausman can reply that he is justified in considering only this mainstream project just because the latter accounts for most of the current output of the economics academy. But this is not good enough for his argument. Hausman is explicitly questioning the reasons for designating the project with which I and others are involved, as realist. His chosen strategy turns on showing that debates and discussions, etc., to which I am party in economics do not involve the postulation of new unobservable entities, and so forth. But for this to work his analysis must cover all debates, etc., to which I am a party, mainstream or otherwise. And for this, the focus must be a far wider and richer conception than that which mainly turns on preferences and expectations.

Certainly Hausman could not simultaneously maintain both that the realist aspect of the project to which I have contributed is not distinctive in the context of modern economics, and also that because it is so distinctive this aspect should not be considered a part of modern economics.

Hausman and Critical Realism

It must be admitted that Hausman does appear to go further in addressing my own project in the latter part of a later section of his paper headed *Transcendental Realism*. Here he recognizes the questioning of whether or not unobservables exist in the social context cannot reasonably be restricted on an a priori basis, even to discussions of entities and properties; Hausman finally allows the possibility that I may be positing underlying structures and mechanisms, etc., as amongst the proper objects of economic study. Ultimately, though, this section mainly comprises various somewhat erroneous, if often tangential, remarks or assertions, mostly still reflecting Hausman's apparently unquestioning support for the mainstream tradition.

For example, it is suggested that I offer 'economists a false dichotomy. Either they can accept a view of science as exclusively the search for exceptionless regularities among observable events . . . , or they can accept critical realism' (Hausman, 1998, p. 204). Now what is false about this dichotomy? After all critical realism argues that the world is open and structured in complex ways. It is because it is so that event regularities whether strict or partial (i.e., 'demi-regularities') *can* be brought about under certain conditions. Critical realism thus entertains a priori the possibility of event regularities of varying degrees of strictness; it all depends. Mainstream (deductivist) economic modelling, in contrast, requires that strict event regularities (including those covered by well defined probabilistic laws) are ubiquitous. So the choice, the dichotomy, is indeed between science being or not being 'exclusively the search for exceptionless regularities among observable events', between the reductionist claims of deductivism and non-reductionist claims of critical realism.

Hausman supposes that the thesis in question, that economists should abandon deductivism (as a universalising claim) for critical realism, somehow follows from a 'controversial metaphysics', which, 1) precisely maps distinctions in metaphysical categories onto those between observable and unobservables, and, 2) supposes that experiences and aspects of mechanisms cannot themselves be events. I am not sure why Hausman supposes that either set of claims is or would be defended. I personally have never entertained either and would not wish to (although I do not deny, of course, that ex posteriori social structures, powers, mechanisms, processes and tendencies, etc., are found to be in large part unobservable).

Hausman further asks: 'What is gained by assimilating questions concerning the status of, for example, social norms (presumably amongst other social structures) to questions concerning the existence of electrons?' (1998, p. 205). But as I have already indicated there is no assimilating going on; these just are the same sort of questions. Both reflect the postulating of the existence of some disputed kind of entity; and both sets of postulations require investigating. Hausman's purpose in all this seems to be a misguided attempt to reduce all aspects of all structures and mechanisms to the level of 'everyday commonsense understanding' as if thereby all aspects can somehow be treated as free of realist or anti-realist controversy.

Hausman also includes a brief discussion of firms and the supposed 'law of diminishing returns' intended to demonstrate that 'Lawson's emphasis on realism distracts attention from the real issues' (1998, p. 205). But in this Hausman appears to suppose that my critique of economics, if applied to discussions of the firm and 'returns to inputs', would amount to little more than a suggestion that more variables or factors should be included in the analysis (' . . . the law of diminishing returns . . . captures only one factor that generates the complicated phenomena observed. One does not have to be a critical realist to recognize this crucial point . . . ' (Hausman, 1998, p. 205)). The fact that I am arguing that the social world, *including firms*, is in part constituted by intrinsically dynamic (and mostly unobservable) highly internally related structures of powers or capacities, etc., irreducible to any actual realization appears to be less than fully appreciated.

Hausman also makes reference to other 'fundamental "principles" of economics' (1998, p. 211) that I do not have space here to discuss. But my general observation is that in all such examples, like that of the supposed law of diminishing returns just noted, the constituents of economics are being too uncritically presumed. Indeed, from my own perspective a most striking feature of the contribution of Hausman (and of various other 'methodologists') is a failure to recognize the limited relevance to the social realm of 'principles' of the sort identified. And this failure can be explained, it seems to me, only by a continuing refusal to question the relevance of the whole mainstream (deductivist) tradition; to ignore, in particular, its continuing practice of ontological neglect.

Concluding Remarks

Hausman's recent (1998) *Economics and Philosophy* article contains a number of claims to the effect that certain questions or issues are, or are not, relevant to economics. It seems that behind each such claim is a presumption that, in order to qualify as relevant or 'pressing', a question or issue must currently be a focus of attention or a topic of debate amongst (mainstream) economists. This line of reasoning appears itself to be underpinned by a more general presumption that economics reduces to what most economists, including economic methodologists, currently do.

My own rather different starting point has been the (widely recognized) phenomenon that modern economics mostly fails to illuminate the world in which we live and, indeed, is in a state of disarray, coupled with a conviction that we ought to do something about it, and specifically to seek to replace, or at least supplement, dominant strategies with others that are rather more capable of being explanatorily successful and useful. Addressing such matters seems to me to be as relevant or pressing as any issue facing modern economics.

Central to my project, then, has been the need to identify the cause of the discipline's failings. And I have certainly found that the problems turn not just on matters of current concern to modern economists, including methodologists, but at least as much on matters for which far too little concern is shown. Specifically, I have argued that the problems of the modern discipline relate fundamentally to ontological neglect. In consequence, my own endeavour to help improve things has involved explicit and sustained ontological elaboration focusing on implications for explanatory conduct in the social realm. Others have contributed in similar fashion.

The result is a project that has been not only more explicitly and systematically oriented to ontological investigation in economics than most other projects, being concerned indeed to elaborate a social metaphysics for social science, but also found to sustain a conception of social life that is far richer, that contains significantly more 'depth', than most competing conceptions in economics.

The conception defended is likely to be contentious, of course. And, whatever the worth of this project and its results, it is the case that its contribution will always be practically conditioned, fallible, partial and likely transient. But whatever else might or might not be claimed of the project I have been discussing, I think it clearly *is* the case that if we question what is distinctive about it in the context of modern economics, a realist orientation, in the senses indicated throughout, has got more than a little to do with it.

References

Fleetwood, S. 1999. *Critical Realism in Economics: Development and Debate*. Routledge.
Hausman, D. M. 1992. *The Inexact and Separate Science of Economics*. Cambridge University Press.
Hausman, D. M. 1998. 'Problems with realism in economics'. *Economics and Philosophy*, 14:185–213.
Lawson, Tony. 1997a. *Economics and Reality*. Routledge.
Lawson, Tony. 1997b. 'Economics as a distinct social science? On the nature, scope and method of economics'. *Economie Appliquée*, L:5–35.

TWENTY-FIVE

Feminism and Economics

Julie A. Nelson

Julie A. Nelson (1956–) is currently a Senior Research Associate with the Global Development and Environment Institute at Tufts University. She received her Ph.D. in Economics from the University of Wisconsin–Madison. Her work has included methodological reflections on feminist economics and empirical research concerning families and household consumption. Nelson is currently an Associate Editor of *Feminist Economics.* Her most recent book is *Economics for Humans.*

An article in *The Chronicle of Higher Education* of June 30, 1993, reported, "Two decades after it began redefining debates" in many other disciplines, "feminist thinking seems suddenly to have arrived in economics." Many economists, of course, did not happen to be in the station when this train arrived, belated as it might be. Many who might have heard rumor of its coming have not yet learned just what arguments are involved or what it promises for the refinement of the profession. The purpose of this essay is to provide a low-cost way of gaining some familiarity.[1]

Most people associate feminism with a political program, which of course it includes. While there are now many varieties of feminism, they all share a concern with remedying the disadvantages historically born by women. Such a concern has been manifested within the discipline of economics in the form of efforts to encourage the advancement of women within the profession (for example, by the Committee on the Status of Women in the Economic Profession) and sometimes by the application of economic analysis for feminist ends. Less familiar to many economists, however, are the implications for economics of recent feminist theorizing about sexism and

---

Journal of Economic Perspectives, vol. 9 (Spring 1995): 131–48. Reprinted with the permission of the American Economic Association.

I particularly thank Robin Bartlett, Joan Combs Durso, Marianne Ferber, Nancy Folbre, Gloria Helfand, Linda Lucas, Constance Newman, Jean Shackelford, Carl Shapiro, Myra Strober, and Timothy Taylor for helpful suggestions.

science. Feminist scholars have documented how beliefs about gender – that is, beliefs about the characteristics and social roles of men and women – have been important on an intellectual as well as social plane. Recent feminist theory leads to questioning of many basic assumptions and values that undergird current economic practice.

Feminist theory raises questions about the adequacy of economic practice not because economics is in general too objective, but because it is not objective enough. Various value-laden and partial – and, in particular, masculine-gendered – perspectives on subject, model, method, and pedagogy have heretofore been mistakenly perceived as value free and impartial in economics, as in other scientific disciplines. Traditionally, male activities have taken center stage as subject matter, while models and methods have reflected a historically and psychologically masculine pattern of valuing autonomy and detachment over dependence and connection.

The alternative suggested here is not, however, a "feminine" economics in which masculine biases are replaced by feminine ones, nor a "female" economics in which economics by or about women is done differently than economics by or about men. The alternative described in this article is an improvement of all of economics, whether done by female or male practitioners.

Gender and Disciplinary Values

If one believes that the current definitions and methods of economics come from outside of human communities – perhaps mandated by divine intervention, or descending via a Friedmanesque helicopter drop – then of course the idea that such standards could be gender-biased will seem nonsensical. But if we allow that economic practice is human practice, developed and refined within human communities, then the possibility must be admitted that human limitations, interests, and perceptual biases will have effects on the culture of economics. The feminist analysis of economics that will be discussed here starts from the premise that economics, like any science, is socially constructed. Social constructionism should not be mistaken for a claim that "anything goes" or that there are no standards of truth or reliability. It simply recognizes that such standards are determined from within a particular scientific community, not from without.

How, then, might gender influence economics? While women were historically excluded from the economics community, some caution is recommended in moving from the observation of women's exclusion to conclusions about the mechanisms by which gender biases take root. It is necessary to clarify here that feminist scholars make a subtle but important distinction between *sex* and *gender. Sex*, as the term is generally used in feminist scholarship, refers to biological differences between males and females. *Gender*, on the other hand, refers to the associations, stereotypes, and social patterns that a culture constructs on the basis of actual or perceived differences between men and women. Women's lesser average brain weight than men, for example, is a biological characteristic. The nineteenth-century interpretation of this fact as implying that women are therefore less rational is an example of a social belief, that is, a construction of gender (Bleier, 1986).

Most feminist scholars see masculine bias in science as primarily an issue of gender, not of sex. The entrance of more women into scientific disciplines is seen as contributing to the transformation of the disciplines, not because women "bring something different" to the fields by virtue of femaleness, but rather because the illumination of gender biases at the level of the social structure of science makes gender biases at other levels more visible as well.[2] To say that "contemporary economics is masculine," then, is to say that it reflects social beliefs about masculinity, not that it reflects the maleness of its traditional practitioners (Keller, 1986). To say that a less masculine-biased economics would be more adequate is to say that social beliefs about economics must change and that economics would be enriched by a diversity of practitioners, not that economics must be practiced by eunuchs or neuters.

The analysis of links between modern western social beliefs about gender and about science was the accomplishment of groundbreaking work by feminist scholars starting in the 1980s (Bordo, 1987; Harding, 1986; Keller, 1985; Merchant, 1980). Objectivity, separation, logical consistency, individual accomplishment, mathematics, abstraction, lack of emotion, and science itself have long been culturally associated with rigor, hardness – and masculinity. At the same time, subjectivity, connection, "intuitive" understanding, cooperation, qualitative analysis, concreteness, emotion, and nature have often been associated with weakness, softness – and femininity. Such associations were sometimes explicit in the language used by the early scientists to define their endeavor. Henry Oldenburg, an early Secretary of the British Royal Society, stated that the intent of the Society was to "raise a masculine Philosophy . . . whereby the Mind of Man may be ennobled with the knowledge of Solid Truths" (Keller, 1985, p. 52).

Simple recognition that the characteristics most highly valued in economics have a particularly masculine gender association does not, however, suggest a unique response for scholars concerned with the quality of economic practice. One response might be to endorse this association so that we can go on doing as we have always done. If this is masculine economics, so be it. The only alternative to masculine economics, our usual way of thinking about gender tells us, would be emasculated, impotent economics.

Another response might be to turn the tables and seek to replace hard, objective, active, androcentric economics with soft, subjective, passive gynocentric economics. One might focus on cooperation, for example, instead of competition and eschew all quantitative methods in favor of qualitative ones. While this might be appealing to those who consider modern economics to be responsible for all the ills in the world, such a response merely trades one set of biases for another.

A third response, particularly associated with the intellectual currents of postmodernism, might be to "deconstruct" the dualisms on which modern definitions of economics depend. In deconstructionist theory, all human projects are simply texts or discourses to which techniques of literacy criticism can be applied. In this view, neither the distinction science/nonscience nor masculine/feminine reflects any nonlinguistic underlying reality. This approach, however, yields little guidance about how to judge the quality of scientific practice.[3]

A fourth approach is adopted in this article. It does not require endorsing one side or the other of the masculine/feminine dualism, nor forgoing evaluation. The key to this approach lies in an unlinking of our judgments about value – that is, about what is meritorious or less meritorious in economic practice – from our perceptions of gender.

The notion that masculine economics is "good" economics depends on a general cultural association of masculinity with superiority and femininity with inferiority, or, in other words, a mental linking of value (superior/inferior) and gender (masculine/feminine) dualisms. Any reader who might question the asymmetry of this linking, preferring, perhaps, to think of gender differences in terms of a more benign complementarity, should ponder some of the more obvious manifestations of asymmetry in the social domain. Rough "tomboy" girls are socially acceptable and even praised, but woe to the gentle-natured boy who is labeled a "sissy"; a woman may wear pants, but a man may not wear a skirt. The sexist association of femininity with lesser worth implicit in such judgments, it should be noted, is not a matter of isolated personal beliefs but rather a matter of cultural and even cognitive habit.

Research on human cognition suggests that dualisms such as superior/inferior and masculine/feminine play an essential role in structuring our understanding (Lakoff and Johnson, 1980; Nelson, forthcoming, ch. 1).

Human cognition is not limited to such simple two-way associations, however. Consider the different interpretations we can make if we think of gender and value, instead of as marking out the same space, as operating in orthogonal dimensions. Then we can think of there being both valuable and harmful aspects to qualifies culturally associated with masculinity, as well as both valuable and harmful aspects to traits associated with femininity (Nelson, 1992).

Consider, for example, the idea that a "hard" economics is clearly preferable to a "soft" economics. This judgment relies on an association of hardness with valuable, masculine-associated strength, and softness with inferior, feminine-associated weakness. However, hardness may also mean rigidity, just as softness may also imply flexibility. A pursuit of masculine hardness that spurns all association with femininity (and hence with flexibility) can lead to rigidity, just as surely as a pursuit of feminine softness (without corresponding strength) leads to weakness. There is no benefit to "specialization" on the side of one gender: neither rigidity nor weakness, the two extremes of hardness and softness, is desirable. There is benefit, however, from exploiting complementarity. Strength tempered with flexibility would yield a balanced and resilient economics. This is just one abstract example of how new thinking about gender could change how we think about discipline; many more concrete examples follow.

Four Aspects of Economics

Applying the feminist scholarship on science to economics suggests that the criteria by which we judge "good economics" have been biased, and that the use of less-biased criteria of evaluation would lead to a more adequate practice. Consider the biases that arise in four different aspects of economics: model, methods, topics, and pedagogy. While critiques and new directions concerning the subject matter of economics and teaching may be familiar to some economists, the more subtle areas of model and method will be discussed first since these have implications for the broadest range of economic practice.

Economic Models

At the center of mainstream economic modeling is the character of the rational, autonomous, self-interested agent, successfully making optimizing choices subject to exogenously imposed constraints. In adopting this conception of human nature, economists have carried out the suggestion of

Thomas Hobbes (as cited in Benhabib, 1987), who wrote, “Let us consider men . . . as if but even now sprung out of the earth, and suddenly, like mushrooms, come to full maturity, without all kind of engagement to each other.” Economic man springs up fully formed, with preferences fully developed, and is fully active and self-contained (England, 1993). As in our Robinson Crusoe stories, he has no childhood or old age, no dependence on anyone, and no responsibility for anyone but himself. The environment has no effect on him, but rather is merely the passive material over which his rationality has play. Economic man interacts in society without being influenced by society: his mode of interaction is through an ideal market in which prices form the only, and only necessary, form of communication.

This is not to say that all practicing economists believe that humans are no more than *homo economicus* (though there are a few true believers), but only that this model of human behavior is perceived as being the most useful and most rigorously objective starting point for economic analysis. Consider, however, the gendered biases implicit in taking the “mushroom man” as representative of what is important about human beings. Humans do not simply spring out of the earth. Humans are born of women, nurtured and cared for as dependent children and when aged or ill, socialized into family and community groups, and are perpetually dependent on nourishment and a home to sustain life. These aspects of human life, whose neglect is often justified by the argument that they are unimportant, or intellectually uninteresting, or merely natural, are, not just coincidentally, the areas of life thought of as “women’s work.”

One must be careful here, again, to draw a distinction between analysis at the level of sex (biological distinction) and analysis at the level of gender (social beliefs). An interpretation that some might draw from the above contrast might be that next to *homo economicus* to describe men’s autonomous, self-interested behavior, we need a *femina economica* to describe women’s connected, other-oriented behavior. Such an endorsement of separate spheres for men and women is, however, quite opposed to a feminist analysis that sees the gender distinctions as socially constructed rather than biologically determined. *Homo economicus* may not be a good description of women, but neither is he a good description of men. Both the autonomous, rational, detached, masculine projection and the dependent, emotional, connected, feminine one are equally mythical and distorting. Men’s traditional facade of autonomy has always been propped up by the background work of mothers and wives; to believe that women are passive requires turning a blind eye to the activity of women’s lives. What is needed is a conception of behavior that does not confuse gender with judgments about value, nor confuse gender with sex. What is needed is a conception of *human* behavior that can encompass both autonomy and dependence, individuation and relation, reason and emotion, as they are manifested in economic agents of either sex.

Feminists need not reinvent the wheel while looking for ways of building more satisfactory models. One example of a richer model of human behavior that is probably familiar to many economists is George Akerlof and Janet Yellen's (1988) theory of efficiency wages as based on fairness. In their model, agents are not hyperrational, isolated monads, but rather human beings capable of "emotions such as 'concern for fairness'" or jealousy (p. 45) and very concerned with their sphere of personal connections. As they point out, the idea that workers' concern with fairness affects their job performance is in fact borne out by empirical studies done by psychologists guided by equity theory and sociologists guided by social exchange theory. In suggesting that wages may be influenced by fairness considerations, rather than purely by market forces, such a model contributes toward explaining the persistence of non-market-clearing wages and the existence of unemployment.

Similar analysis has suggested that notions of fairness play an important role in the setting of prices in product markets (Kahneman, Knetsch, and Thaler, 1986). Lee Levin's (1995) theory of investment also borrows freely from psychology and sociology to gain insight into economic phenomenon. Levin suggests that Keynes' notion of animal spirits can be fleshed out using theories of convention, rumor, social comparison, fad, cognitive dissonance, and contagion theory borrowed from these other disciplines. Nancy Folbre (1994a), Amartya Sen (1977) and Robert Frank (1988) are economists who have also explored richer models of human economic behavior, both individual and collective. Readers may think of other examples. A degree of care must be maintained, of course, in moving away from the simple rational-choice model or borrowing from other disciplines: overthrowing a model of autonomous choice only to end up with, for example, a model of pure social determinism would lead to no great improvement. But feminist analysis suggests that the current neglect of social and emotional dimensions of human behavior should be considered a serious limitation, rather than a sign of rigor.

The question of economic models overlaps with the question of how economics is to be defined as a discipline. As Akerlof and Yellen's (1988) model explains a particular macroeconomic phenomenon in an empirically supported way, it would seem to clearly qualify as an economic model. Yet some see economics as *defined by* the *homo economicus* model. For them, models like that of Akerlof and Yellen fail to qualify, being too "soft" or "too messy," or perhaps "too sociological." Gary Becker (1976, p. 5), for example, has argued that it is the model of individual choice in markets that is the distinguishing characteristic of economics. Robert Lucas (1987, p. 108) has stated that the assumptions of rational choice modeling provide "the only 'engine of truth' we have in economics." The feminist analysis suggests that Becker's and Lucas' approaches are not, as they are often taken, statements of demand for high rigor, but rather are demands that androcentric biases be indulged.

Such a definition of economics according to (a restrictive) model, rather than subject matter, has been an effective rhetorical strategy for cutting off alternative views (Strassmann, 1993). One might take the growth and acceptance of much of the new classical macroeconomics modeling program, protected by Lucas' definition of the discipline, as a case in point. But such a strategy can retain its effectiveness only so long as the association of masculinity with high value has emotional and cognitive power. The feminist analysis suggests that there should not be just one economic model, but rather many economic models, depending on the usefulness of various modeling techniques in the various applications. Many of these models will still emphasize individual choice and purposive behavior, but some will not. To argue that economists should continue to specialize in a single specific type of model, because that is how we have been trained, is to argue that sunk costs should play a role in determining current profit-maximizing choices – a fallacy we usually try to debunk in our undergraduate students' sophomore year. An efficient business certainly would not allow an employee to continue practicing a skill that yields low returns because of an oversupply or a changing market, just because the skill was difficult and time consuming to acquire.

While feminist economics does not impose feminist policy conclusions on economic research, it can be noted that such a broadening of economic modeling opens new opportunities in the analysis of labor market discrimination. Within a model of rational, autonomous individual behavior and perfectly clearing markets, women's lower earnings and exclusion from certain professions tan be explained only by appeal to extra-market sources, such as women's career and education decisions or the amount of effort women put forth (for a review see Bergmann, 1986). Employer discrimination cannot persist in competitive markets, goes Becker's story, since discrimination is a taste that is costly to indulge. Discriminators will hence be outcompeted by firms that make profit-maximizing choices. Comparable worth is a political rather than an economic issue, it is sometimes said, since the idea that occupations held largely by women could be systematically underpaid is in violation of the thesis that wages are determined by market forces. The influence of such positions is not based in the empirical support they have garnered, however: the strength of their appeal to economists lies only in their consistency with the narrow choice-theoretic model. Broader models that include the social and emotional factors ignored in standard neoclassical analysis make room for discrimination as a potential issue.

If employers are themselves subject to widespread and systematic social pressures, for example, *nondiscrimination* might be a taste that is costly to indulge. Employers may meet not only with rebellion from their other workers but with ostracism from their peers and perhaps even from their friends and family when they violate widespread gender and racial norms in hiring or compensation (Strober and Arnold, 1987). If wages reflect perceptions of fairness, as Akerlof and Yellen (1988) have argued, then perceptions of the relative worth of men's and women's work is quite relevant to wage determination. If, as feminists argue, certain traits and jobs traditionally associated with women have been systematically undervalued, it may be perceived as fair to pay less for these skills (England, 1992).

The feminist insight into economic modeling does not prescribe in advance that injustice will be found in every study of the labor market. It does require, however, that we not dismiss the possibility that wages may depend on factors beyond marginal products simply because the models we use are blinded by their own assumptions.

Economic Methods

While models of individual rational choice could conceivably be expressed and analyzed in a purely verbal manner, it seems almost a tautology to say that in the discipline of economics, quality in method is identified primarily with mathematical rigor. Strict adherence to rules of logic and mathematics, formalization in the presentation of assumptions and models, sophistication in the application of econometric techniques – these are the factors, in many people's minds, that set economics apart from "softer" fields like sociology or political science. Use of formal and mathematical methods (particularly in the form of constrained maximization) is also often presumed to assure the objectivity of economic results. Abstract and highly formalized analysis is often valued over concrete and detailed empirical work, for the logical purity of its proofs and for its context-free generality. While good writing and verbal analysis do not go entirely unrewarded, they are usually considered to be largely auxiliary to the real analysis.

Feminist scholarship suggests that such narrow views of knowledge and rationality are holdovers from a crisis about masculinity during the early years of the development of modern science, particularly manifested in the ascendancy of Cartesian philosophy (Bordo, 1987; Easlea, 1980). Far from protecting economics against biases, such a concentration on toughness and detachment hog-ties our methods of analysis. Emphasis on being hard, logical, scientific, and precise has served a valuable purpose, it is true, in guarding against analysis that is weak, illogical, unscientific, and vague. But if these are the *only* virtues we value in our practice, we are easy prey to other vices.

Emphasis on masculine hardness without flexibility can, as discussed above, turn into rigidity. Emphasis on logic, without sufficient attention to grasping the big picture, can lead to empty, out-of-touch exercises in pointless deduction. Scientific progress without attention to human values can serve inhuman ends. Arguments that have given up all richness for the sake of precision end up being very thin. Including both masculine- and feminine-identified positive qualities, on the other hand, makes possible a practice that is flexible, attentive to context, humanistic, and rich as well as strong, logical, scientific, and precise.

Feminist economists are not the only economists to voice dissatisfaction with the narrow strictures put on knowledge seeking in economics, that as a consequence leave economists inadequately educated and inadequately practiced in skills of richer and more substantive analysis. While feminist theorists offer a unique explanation for the psychological and social tenacity of the Cartesian view (and link it to failures in areas of model, topics, and pedagogy as well), feminist economists hardly need to start from scratch in envisioning a more adequate methodological toolbox. Donald McCloskey, for example, has written extensively on the possibilities for improvement of rhetorical standards within the profession. McCloskey (1993) argues that feminine-associated argumentation by metaphor and story must be given equal scientific prestige with masculine-associated argumentation by fact and logic. As a practical matter, he has even published a short guide to improved writing (McCloskey, 1987).

The Commission on Graduate Education in Economics (COGEE) of the American Economic Association recently expressed concern about overemphasis on context-free analysis. Its report noted a fear that economics "graduate programs may be turning out a generation with too many *idiot savants* skilled in technique but innocent of real economic issues" (Krueger, et al., 1991, pp. 1044–45). While this report set up the problem as one of an imbalance between (mathematical, technical) methods on the one hand

and substance on the other, with the methodological approach itself left unchallenged, such an argument seems to suggest that knowledge of "facts, institutional information, data, real-world issues, applications and policy problems" (p. 1046) occurs by direct absorption. Yet careful and systematic seeking out of information and good nonformal reasoning about real economic issues can only be accomplished through mastery of corresponding skills, such as (to start with) library research methods and techniques of critical reading. If it is recognized that such skills are just as valid, and just as teachable, as formal and abstract techniques, then the problem indeed includes the issue of balance in methods. As a practical matter, the COGEE report includes a number of specific suggestions about prerequisites, course syllabi, content, assignments, etc. that would begin to move graduate departments somewhat toward educating students to become more competent in analyzing actual economic problems.

Calls for increased attention to the nuts and bolts of empirical work by distinguished economists such as Thomas Mayer (1993) and Lawrence Summers (1991) are also calls for changes in the value system of economics that feminists can join and endorse. Economists tend to be highly skilled in mathematical and statistical theory. However, we generally demonstrate far less skill in other aspects of scientific empirical work like the seeking out of new data sources, the improvement of data collection, responsible data cleaning and quality evaluation, replication, sensitivity testing, proper distinction between statistical and substantive significance, and data archiving (for example, Dewald, Thursby, and Anderson, 1986). Empirical work characterized by such a continual refinement of abstract theory, accompanied by an egregious neglect of concrete detail, has been described by Mayer (1993, p. 132) as "driving a Mercedes down a cow track." While a few journals and funding agencies have in recent years sought to raise professional standards by, for example, requiring sensitivity tests and archiving of data, much more could be done by journal editors, funding bodies, and dissertation advisors to encourage more efforts in these directions. Graduate studies committees could do more to provide the sorts of course work and experience in which good technique could be learned. The feminist critique suggests that there may be much to be gained by decreased use of the technique of detached "musing" (Bergmann, 1987a) and increased use of the technique of "hobnobbing with one's data" (Strober, 1987).

Value judgments attached to "hard" versus "soft" data also deserve reexamination. Economists' skepticism about asking people about the motives behind their behavior is so strong that Alan Blinder (1991) devoted a full section in a recent piece on price stickiness simply to justify the use of such interview survey data. Judged by a standard of Cartesian "proof," such evidence may be inadmissible. But judged by a broader and more practical standard of learning about economic functioning, such data can be seen to potentially contribute important information. A recent conference of the International Association for Feminist Economics, for example, included presentations by a historian and a sociologist on the techniques of doing oral history studies. Economists who overcome their prejudice in this area may be surprised at the sophistication in technique and the attention given to issues of validity and replicability demonstrated by those trained in such "soft" and qualitative methods.

Personal experience should also not be discounted among ways in which we – consciously or not – gather data. It has been a matter of some ironic comment among feminist economists, for example, how what often really seems to matter in convincing a male colleague of the existence of sex discrimination is not studies with 10,000 "objective" observations, but rather a particular single direct observation: the experience of his own daughter.

The idea that one's personal, "subjective" position and opinions could influence the outcome of one's scientific work is, of course, anathema to those who believe that objectivity in scientific pursuits can be attained only by the cool detachment of the researcher from the subject of study, or that objectivity is assured by an individual's strict adherence to particular methods of inquiry. Such a notion of objectivity is considered in the feminist analysis (as well as in much contemporary philosophy of science) to be one more outgrowth of the Cartesian illusion. Part of the practice of striving for objectivity, in fact, should be an examination of how the things that one believes from one's own experience may influence one's research. Sandra Harding (1995) calls the sort of objectivity in which one recognizes one's standpoint "strong objectivity," as contrasted to "weak objectivity" in which the issue of perspective is kept under wraps. Amartya Sen (1992, p. 1) similarly argues that objectivity begins with "knowledge based on positional observation." The movement from subjective views to (strong) objectivity comes not through a sharp separation of the researcher from the object of study, but rather through a connection of the researcher to a larger critical community. According to feminist philosopher Helen Longino (1990, p. 79), "The objectivity of individuals . . . consists in their participation in the collective give-and-take of critical discussion and not in some special relation (of detachment, hardheadedness) they may bear to their observations." While concern with the reliability of results is still of prime importance, the criteria that guide research are internal to the community of researchers, not external. Formalization, rather than reflecting the height of objectivity, is simply seen as one tool in the toolbox. In the words of Knut Wicksell (quoted in Georgescu-Roegen, 1971, p. 341), the role of logic and abstraction is "to facilitate the argument, clarify the results, and so guard against possible faults of reasoning – that is all."

Economic Topics

A prototypical economic article uses an economic model, economic methods, and is on an economic topic. Beneficial expansions in the first two areas were discussed above; the last consideration is, however, crucial as well. A broad definition of the core topic of economics to which most economists might agree is that of markets. Economics is often defined as the study of processes by which things – goods, services, financial assets – are exchanged. By this definition, most of the traditional nonmarket activities of women – care of the home, children, sick and elderly relatives, and so on – have been considered "noneconomic" and therefore inappropriate subjects for economic research. Families, in fact, often seem to disappear entirely from the world of economists. Consider this textbook discussion: "The unit of analysis in economics is the individual . . . [although] individuals group together to form collective organizations such as corporations, labor unions, and governments" (Gwartney, Stroup, and Clark, 1985, p. 3).[4] Families are, apparently, too unimportant to mention.

The most notable exceptions to this neglect, of course, come from Gary Becker and the other so-called "new home economists" and from more recent developments in the application of game theory to the family. The existence of these literatures is something of a double-edged sword to feminists. On one hand, they do bring into mainstream journals some discussion of family issues. However, they conform strictly to the narrow standards of method and model discussed earlier, and it is probably only by these criteria that they hang on to their "economic" credentials. Moreover, it is troublesome to feminists that this work has often assumed or endorsed traditional expectations about the sexes. While Becker has indeed developed models of family interactions, he, as Barbara Bergmann (1987b, pp. 132–33) has put it, "explains, justifies, and even glorifies role differentiation by sex . . . to say that the 'new home economists' are not feminist in their orientation would be as much of an understatement as to say that Bengal tigers are not vegetarians."

While some debate has gone on about Becker's models, and the possibility of using individualistic rational-choice models for feminist ends, such battles take place at the margins of economics. The expectation on the part of many economists that feminists will or should concentrate on debating Becker may primarily serve as a handy way to avoid engagement with feminist critique. Such a view limits feminist critique to a field that is perhaps safely distant from one's own.

While families are "economic" to Becker to the extent that they can be modeled in terms of choices and markets, families have traditionally been "economic" to women in a much more direct sense. Many women's economic security historically was, and to some extent still is, far less dependent on their own earnings than on whether or not they "marry well." Further, while economists and census takers have waffled back and forth on whether unpaid housekeeping should be classified as leisure or work (Folbre, 1991), the women scrubbing the sink rarely entertained any doubt.

Drawing the distinction about what is "economic" and what is not at the household door leads increasingly to odd dead ends and bifurcations in economic analysis. Why should childcare, elder care, and care of the sick be "economic" when provided by markets (or sometimes government), but not worthy of study by economists when done in private homes? Rather than using marketization as the criterion for demarcating economics – or using the rational choice model, as discussed above – a broader definition of economics as concerned with "provisioning" could delineate a subject matter without using sexist assumptions about what is and what is not important (Nelson, 1993b).

Adam Smith, for example, defined economics not as simply about choice and exchange, but also as about the production and distribution of all of the "necessaries and conveniences of life," placing emphasis on the things that human beings need to survive and flourish. These things may include activities such as meaningful work, as well as goods and services such as food and health care. While some goods and services may be freely chosen by adult individuals acting in markets, many are provided to individuals by their parents during childhood or by other family members. They can also be provided as gifts or through community or governmental programs. The distribution of many "necessaries and conveniences" is also strongly influenced by tradition and coercion.

Such a definition of economics as concerned with the realm of "provisioning" breaks down the usual distinction between "economic" (primarily market-oriented) activities and policies on the one hand, and familial or social activities and policies on the other. The absence of entries for household production in the National Income and Product Accounts illustrates the way in which such a bifurcation has structured economic analysis, and the concern of many feminists about this neglect is relatively well known (for example, Waring, 1988). The priority to be given to a project of inclusion of household activities in gross domestic product is, however, actually somewhat controversial among feminists, with some arguing that increased emphasis on housework would only serve to glorify the homemaker role, and with many concerned that monetary figures would be downward biased due to the low value currently given to activities like childcare (Folbre, 1994b). Questions can also be raised about the importance to be given to the GDP numbers themselves. Feminists may join with others in criticizing the methodological reductionism of frequently using such a crude measure of market and governmental economic activity as a yardstick for economic welfare. Multidimensional measures, which might include measures of distribution and sustainability, and measures of human outcomes such as educational attainment and health (Nussbaum and Sen, 1993) would form a more adequate basis for economic analysis and national policy-making and evaluation. Accounting for the division of labor and of goods within the household, for example, has been particularly important in feminist work on development economics, economies in transition, and economies undergoing structural adjustment (Sen, 1985; Bakker, 1994).

Less well known than the GDP critique is the concern of many feminists about adequate attention to investment in children and the question of who bears the costs of such investment (Folbre, 1994a). Programs to improve child nutrition or preschool and primary education, for example, are usually thought of as "social" programs, merely frosting on the fiscal cake, rather than as economic programs, designed to advance investment in human capital. Programs to increase the quality of paid childcare arrangements are often thought of as consumption goods for parents, rather than as investments in children and in necessary infrastructure for parental (and particularly, given stereotyped patterns of work distribution within families, mothers') participation in the life of the community. Standard economic analysis and pedagogy tend to reinforce such trivialization: recall that the mythical economic "mushroom man" springs up without any need for provisioning by others in his youth, and recall that the actual work by women in historically providing such direct provisioning does not count as "economic."

Consider, as a specific example, the treatment of the subject of human capital in a standard labor economics textbook. It does not begin with a focus on nutrition, socialization, and informal and formal education of children within families and public schools, but rather with the college choice decision of the young adult (Ehrenberg and Smith, 1994). While there is, of course, some pedagogical benefit in designing textbook examples to appeal to the immediate interests of the students, there seems to be no pedagogical reason to focus so narrowly on higher education. When I teach this subject, I find it helpful to add readings about earlier human capital development and to ask students to reflect back on the creation of their abilities and aspirations, from the very start of their lives.

The question of topics cannot be unlinked from the earlier questions raised about models and methods. Consider how biases in models and methods have distorted the development of one research program that from the start has been concerned with household issues: the economics literature on "household equivalence scales." Such scales adjust measures of household income for differences in household size and composition. These scales are in daily policy use in, for example, setting equitable levels of social benefits across households of different size, as well as being frequently used by researchers in studies of income distribution.

The economic literature on household equivalence scales has, however, moved further and further away from questions of policy relevance (Nelson, 1993a). First, while policy applications are often concerned centrally with the welfare of children (for example, in setting levels for Aid to Families with Dependent Children), forcing the question into a utility-theoretic framework (in which the scale is interpreted as a ratio of expenditure functions) has led to household welfare being generally modeled as the welfare of the *adults*. Some of the most highly regarded models imply that adults will, in fact, substitute *away* from goods consumed largely by children when they are present. Second, while early empirical estimation of equivalence scales depended largely on prescriptive budget studies that listed how much each type of household would "require" for food, rent, and so on, more recent empirical practice has been characterized by greater subtlety and sophistication. Unfortunately, the estimation of scales using large-scale demand system regressions informed by specific utility-theoretic models, which was for a period the norm, has for the most part turned out to be fundamentally underidentified (since the same demand equations may be consistent with any number of expenditure functions).

While many economists have followed these trends to greater reliance on choice theory and advanced econometrics, a recent piece by Trudi Renwick and Barbara Bergmann (1993) gives one illustration of what can be accomplished when the focus stays closer to the policy question, with less allegiance to particular models and methods. Renwick and Bergmann's formulation of "basic needs" budgets for households with different compositions follow the earlier prescriptive budgets approach, and is updated for changing times by the addition of childcare expenditures. While technically unsophisticated, one may actually learn more about costs from such direct (if admittedly approximate and prescriptive) evidence, than from more technically sophisticated but unfocused and indirect techniques.

Economic Pedagogy

A discipline of economics defined around a formal rational-choice model, with perhaps a few facts delivered on the side, can perhaps be adequately perpetuated by a style of teaching that focuses purely on the transfer of preset knowledge. If economics is defined more broadly, however, such an approach may not be adequate. Fostering the ability to think critically, analytically, and creatively about economic issues requires a different pedagogy. While such a way of thinking may be as teachable as current methods, as argued above, it may not be necessarily as *easily* teachable, nor teachable by the same teaching methods – nor perhaps even to exactly the same students.

Feminist economists suggest that not only the content of economics courses, but also the teaching style used could undergo a beneficial transformation (Strober, 1987; Bergmann, 1987a; Bartlett and Feiner, 1992; Shackelford, 1992). Some emphasize the use of experimental learning and laboratory sessions in which students work with simulations, collect their own interview data, and/or analyze data, to give students more chance to "do economics" and work out the answers to questions (Bartlett and King, 1990). Some suggest that feminist pedagogy requires a different relationship between the professor and students, with less distance and more dialogue between the professor and students, and also among students. Some suggest that explicit attention be paid to the affective aspects of learning (Strober, 1987).

Feminists are, of course, not the only educators interested in more interactive and cooperative learning; to many educators, this is just "good pedagogy," as demonstrated by studies of how students actually learn. Feminist theorists are more likely, however, to see the resistance to pedagogical reform as being rooted in general cultural associations of gender and value. These pedagogical insights are intended to apply across the curriculum, not just to "women and the economy" courses. Active learning techniques could improve the practice of even the more familiar forms of economic analysis. An ability to think critically is arguably as important in making good judgments about the choice of statistical methods or the use of significance tests, for example, as it is in writing an essay on antidiscrimination policies.

One point of interest to many feminists is the way in which economic pedagogy may subtly shape the demographic composition of future economists.

Much has been written about the way in which the "classroom climate" – including instructors' patterns of interaction with men and women students and sex stereotyping in textbooks – may make women less confident about succeeding in particular areas (Hall and Sandler, 1982; Ferber, 1990). The standard androcentric biases in the topics, models and methods of economics may be added to the list of ways in which women students may be subtly influenced to believe that "economics is not for (or about) me." The current emphasis on mathematical technique also leads to self-selection of those students, male and female, who find abstract analysis satisfying but who may be weak in broader analytical thinking, and the self-exclusion of many students who perhaps have fine analytical skills, but see little use for them in economics. Such selection leads to a vicious cycle, in which students and instructors are both heavily invested in the status quo.

Conclusion

Feminist economics, to reiterate, is not female economics, to be practiced only by women, nor feminine economics that uses only soft technique and cooperative models. Feminist scholarship suggests that economics has been made less useful by implicitly reflecting a distorted ideal of masculinity in its models, methods, topics, and pedagogy. Feminist scholars argue that the use of a fuller range of tools to study and teach about a wider territory of economic activity would make economics a more productive discipline for both male and female practitioners.

Many readers may have discovered that they are already doing "feminist economics" in some ways, although they have preferred to think of themselves as just doing "good economics." If one feels a need to defend one's work from the description "feminist," it might be enlightening to ask oneself about the source of this discomfort. Perhaps such defensiveness reflects cultural beliefs about masculinity and femininity, and superiority and inferiority, that could stand some examination.

Notes

1. While isolated feminist challenges to neoclassical theories date back at least to the 1970s (for example, Bell, 1974), this article focuses on the "second revolution" (Coughlin, 1993) that has taken place in just the last few years. This stronger current is exemplified by publications such as Ferber and Nelson (1993), recent sessions on feminist economics at the meetings of the American Economic Association (Bartlett and Feiner, 1992; Shackelford, 1992; Strassmann, 1994; Strober,

1994), and the organization in 1992 of the International Association for Feminist Economics.
2. This is not to ignore the current scholarly (and popular) debates about the degree to which males and females may "think differently" due to genetic or hormonal conditions, only to note that it is not the central issue here. Carol Gilligan's (1982) work, for example, is often quoted in these discussions as having uncovered male/female psychological differences. A more careful reading of her work, however, and of studies which have followed, indicates considerable overlap between men and women in the dimensions studied. What *is* important to the point here, and what has been shown in a number of studies, is the way in which people in U.S. and European cultures tend to mentally *associate* certain characteristics with masculinity or femininity. For a review of these literatures, see Jane Mansbridge (1993), especially notes 32 and 53.
3. While some feminist economists take a thoroughly postmodernist position, and a few even expound a gynocentric view, this essay focuses on those lines of feminist theorizing about economics that seem to have the most adherents, and that, to mainstream economists, should provide the most convincing case for reform.
4. I thank Marianne Ferber for this example.

References

Akerlof, George A., and Janet L. Yellen, "Fairness and Unemployment," *American Economic Review*, May 1988, *78*:2, 44–9.

Bakker, Isabella, ed., *The Strategic Silence: Gender and Economic Policy*. London, Zed Books, 1994.

Bartlett, Robin L., and Susan F. Feiner, "Balancing the Economics Curriculum: Content, Method, and Pedagogy," *American Economic Review*, May 1992, *82*:2, 559–64.

Bartlett, Robin L., and Paul King, "Teaching Economics as a Laboratory Science," *Journal of Economic Education*, Spring 1990, *21*, 181–93.

Becker, Gary S., *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, 1976.

Bell, Carolyn Shaw, "Economics, Sex and Gender," *Social Science Quarterly*, 1974, *55*:3, 615–31.

Benhabib, Seyla, "The Generalized and the Concrete Other: The Kohlberg-Gilligan Controversy and Moral Theory." In Meyers, Diana, and Eva Feder Kittay, eds., *Women and Moral Theory*. Totowa, N.J.: Rowman and Littlefield, 1987, pp. 154–77.

Bergmann, Barbara R., *The Economic Emergence of Women*. New York: Basic Books, 1986.

Bergmann, Barbara R., "'Measurement,' or Finding Things Out in Economics," *Journal of Economic Education*, Spring 1987a, *18*:2, 191–203.

Bergmann, Barbara R., "The Task of a Feminist Economics: A More Equitable Future." In Farnham, Christie, ed., *The Impact of Feminist Research in the Academy*. Bloomington: Indiana University Press, 1987b, pp. 131–47.

Bleier, Ruth, "Sex Differences Research: Science or Belief?" In Bleier, Ruth, ed., *Feminist Approaches to Science*. New York, Pergamon Press, 1986, pp. 147–64.

Blinder Alan S., "Why Are Prices Sticky? Preliminary Results from an Interview Survey," *American Economic Review*, May 1991, *81*:2, 89–96.

Bordo, Susan, *The Flight to Objectivity: Essays on Cartesianism and Culture.* Albany: State University of New York Press, 1987.

Coughlin, Ellen, K., "Feminist Economists vs. 'Economic Man': Questioning a Field's Bedrock Concepts," *Chronicle of Higher Education*, June 30, 1993, *39*:43, A8–A9.

DeWald, William G., Jerry G. Thursby, and Richard G. Anderson, "Replication in Empirical Economics," *American Economic Review*, September 1986, *76*:4, 587–603.

Easlea, Brian, *Witch Hunting, Magic and the New Philosophy: An Introduction to Debates of the Scientific Revolution, 1450–1750.* Atlantic Highlands, N.J.: Humanities Press, 1980.

Ehrenberg, Ronald G., and Robert S. Smith, *Modern Labor Economics: Theory and Public Policy.* 5th ed. New York: Harper Collins Publishers, 1994.

England, Paula, *Comparable Worth: Theories and Evidence.* New York: Aldine de Gruyter, 1992.

England, Paula "The Separative Self: Androcentric Bias in Neoclassical Assumptions." In Ferber, Marianne A., and Julie A. Nelson, eds., *Beyond Economic Man: Feminist Theory and Economics.* Chicago: University of Chicago Press, 1993, pp. 37–53.

Ferber, Marianne A., "Gender and the Study of Economics." In Saunders, Phillip, and William B. Walstad, eds., *The Principles of Economics Course: A Handbook for Instructors.* New York, McGraw-Hill, 1990, pp. 44–60.

Ferber, Marianne A., and Julie A. Nelson, *Beyond Economic Man: Feminist Theory and Economics.* Chicago: University of Chicago Press, 1993.

Folbre, Nancy, "The Unproductive Housewife: Her Evolution in Nineteenth-Century Economic Thought," *Signs: Journal of Women in Culture and Society*, Spring 1991, 16:3, 463–84.

Folbre, Nancy, *Who Pays for the Kids? Gender and the Structures of Constraint.* New York: Routledge, 1994a.

Folbre, Nancy, "Domesticate the Gross Product," *Dollars and Sense*, March/April 1994b, *192*, 7.

Frank, Robert H., *Passions within Reason: The Strategic Role of the Emotions.* New York: W. W. Norton, 1988.

Georgescu-Roegen, Nicholas, *The Entropy Law and the Economic Process.* Cambridge: Harvard University Press, 1971.

Gilligan, Carol, *In a Different Voice: Psychological Theory and Women's Development.* Cambridge: Harvard University Press, 1982.

Gwartney, James D., Richard Stroup, and J. R. Clark, *Essentials of Economics.* New York: Academic Press, 1985.

Hall, Roberta M., and Bernice R. Sandler, *The Classroom Climate: A Chilly One for Women.* Washington D.C.: Association of American Colleges, 1982.

Harding, Sandra, *The Science Question in Feminism.* Ithaca: Cornell University Press, 1986.

Harding, Sandra, "Can Feminist Thought Make Economics More Objective?," *Feminist Economics*, Spring 1995, *I*:1, forthcoming.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, September 1986, *76*:4, 728–41.

Keller, Evelyn Fox, *Reflections on Gender and Science.* New Haven, Conn.: Yale University Press, 1985.

Keller, Evelyn Fox, "How Gender Matters: Or, Why It's So Hard for Us to Count Past Two." In Harding, Jan, ed., *Perspectives on Gender and Science.* London: Palmer Press, 1986, pp. 168–83.

Krueger, Anne O., et al., "Report of the Commission on Graduate Education in Economics," *Journal of Economic Literature*, September 1991, *29*:3, 1035–53.

Lakoff, George, and Mark Johnson, *Metaphors We Live By.* Chicago: University of Chicago Press, 1980.

Levin, Lee B., "Toward a Feminist, Post-Keynesian Theory of Investment: A Consideration of the Socially- and Emotionally-Constituted Nature of Agent Knowledge." In Kuiper, Edith, Jolande Sap, Susan Feiner, Notburga Ott, and Zafiris Tzannatos, eds., *Out of the Margin: Feminist Perspectives on Economic Theory*. London: Routledge, forthcoming, 1995.

Longino, Helen, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press, 1990.

Lucas, Robert E., Jr., *Models of Business Cycles.* Oxford Basil Blackwell, 1987.

Mansbridge, Jane, "Feminism and Democratic Community." In Chapman, John W., and Ian Shapiro, eds., *Democratic Community: NOMOS XXXV*. New York: New York University Press, 1993, pp. 339–95.

Mayer, Thomas, *Truth vs. Precision in Economics.* Brookfield. Vt.: Edward Elgar, 1993.

McCloskey, Donald N., *The Writing of Economics.* New York: Macmillan, 1987.

McCloskey, Donald N., "Some Consequences of a Conjective Economics." In Ferber, Marianne A., and Julie A. Nelson, eds., *Beyond Economic Man: Feminist Theory and Economics.* Chicago: University of Chicago Press, 1993, pp. 69–93.

Merchant, Carolyn, *The Death of Nature: Women, Ecology and the Scientific Revolution.* San Francisco Harper & Row, 1980.

Nelson, Julie A., "Gender, Metaphor and the Definition of Economics," *Economics and Philosophy*, Spring 1992, *8*:1, 103–25.

Nelson, Julie A., "Household Equivalence Scales: Theory vs. Policy?," *Journal of Labor Economics*, July 1993a, *11*:3, 471–93.

Nelson, Julie A., "The Study of Choice or the Study of Provisioning? Gender and the Definition of Economics." In Ferber, Marianne A., and Julie A. Nelson, eds., *Beyond Economic Man: Feminist. Theory and Economics.* Chicago: University of Chicago Press, 1993b, pp. 23–36.

Nelson, Julie A., *Feminism, Objectivity, and Economics.* London: Routledge, forthcoming.

Nussbaum, Martha C., and Amartya Sen, eds., *The Quality of Life.* Oxford Clarendon Press, 1993.

Renwick, Trudi J., and Barbara R. Bergmann, "A Budget-Based Definition of Poverty: With An Application to Single-Parent Families," *Journal of Human Resources*, Winter 1993, *28*:1, 1–24.

Shackelford, Jean, "Feminist Pedagogy: A Means for Bringing Critical Thinking and Creativity to the Economics Classroom," *American Economic Review*, May 1992, *82*:2, 570–76.

Sen, Amartya, "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy and Public Affairs*, Summer 1977, *6*, 317–44.

Sen, Amartya, "Women, Technology, and Sexual Divisions." In United Nation Conference on Trade and Development, *Trade and Development: An UNCTAD Review*, 1985, *6*, pp. 195–223.

Sen Amartya, *Objectivity and Position.* The Lindley Lecture, The University of Kansas, 1992.

Strassmann, Diana L., "Not a Free Market: The Rhetoric of Disciplinary Authority in Economics." In Ferber, Marianne A., and Julie A. Nelson, eds., *Beyond Economic Man: Feminist Theory and Economics.* Chicago: University of Chicago Press, 1993, pp. 54–68.

Strassmann, Diana L., "Feminist Thought and Economics: Or, What do the Visigoths Know?," *American Economic Review*. May 1994, *84*:2, 153–58.

Strober, Myra, "The Scope of Microeconomics: Implications for Economic Education," *Journal of Economic Education*, Spring 1987, *18*, 135–49.

Strober, Myra, "Rethinking Economics Through a Feminist Lens," *American Economic Review*, May 1994, *84*:2, 143–47.

Strober, Myra, and Carolyn L. Arnold, "The Dynamics of Occupational Segregation among Bank Tellers!" In Brown, Clair, and Joseph A. Pechman, eds., *Gender and the Workplace.* Washington D.C.: The Brookings Institution, 1987, pp. 107–48.

Summers, Lawrence H., "The Scientific Illusion In Empirical Macroeconomics," *Scandinavian Journal of Economics*, 1991, *93*:1, 129–48.

Waring, Marilyn, *If Women Counted: A New Feminist Economics.* New York: Harper & Row, 1988.

TWENTY-SIX

Credible Worlds

The Status of Theoretical Models in Economics

Robert Sugden

Robert Sugden (1949–) is a professor of economics at the University of East Anglia. His research uses theoretical, experimental, and philosophical methods to investigate issues in welfare economics, choice under uncertainty, pro-social behavior, the emergence of conventions and norms, economic methodology, and philosophical economics. He is probably best known for developing "regret theory" (with Graham Loomes) and for *The Economics of Rights, Cooperation and Welfare*, one of the first applications of evolutionary game theory to social theory and moral philosophy. Currently, he holds a research fellowship from the Economic and Social Research Council for work on reconciling behavioral and normative economics.

Overview

Using as examples Akerlof's 'market for "lemons"' and Schelling's 'checkerboard' model of racial segregation, this paper asks how economists' abstract theoretical models can explain features of the real world. It argues that such models are not abstractions from, or simplifications of, the real world. They describe counterfactual worlds which the modeller has constructed. The gap between model world and real world can be filled only by inductive inference, and we can have more confidence in such inferences, the more credible the model is as an account of what could have been true.

1. Introduction

I write this paper not as a methodologist or as a philosopher of social science – neither of which I can make any claim to be – but as a theoretical economist. I have spent a considerable part of my life building economic models, and examining the models that other economists have built. I believe that I am

Journal of Economic Methodology, vol. 7 (2000): 1–31. Reprinted by permission of Taylor and Francis, Ltd.

making reasonably good use of my talents in an attempt to understand the social world. I have no fellow-feeling with those economic theorists who, off the record at seminars and conferences, admit that they are only playing a game with other theorists. If their models are not intended seriously, I want to say (and do say when I feel sufficiently combative), why do they expect me to spend my time listening to their expositions? Count me out of the game. At the back of my mind, however, there is a trace of self-doubt. Do the sort of models that I try to build really help us to understand the world? Or am I too just playing a game, without being self-critical enough to admit it?

My starting point is that model-building in economics has serious intent only if it is ultimately directed towards telling us something about the real world. In using the expression 'the real world' – as I shall throughout the paper – I immediately reveal myself as an economic theorist. This expression is standardly used by economic theorists to mark the distinction between the world inside a model and the 'real' world outside it. Theory becomes just a game when theorists work entirely in the world of models. As an analogy, we might think of chess, which was once a model of warfare, but has become a game – a self-contained world with no reference to anything outside itself.

My strategy is to focus on two models – George Akerlof's 'market for lemons', and Thomas Schelling's 'checkerboard city' – which exemplify the kind of model-building to which I aspire. Of course, these are not typical examples of economic models: they represent theory at its best. Nevertheless, at least at first sight, these models have many of the vices that critics attribute to theoretical economics: they are abstract and unrealistic and they lead to no clearly testable hypotheses. It would be easy to caricature them as examples – perhaps unusually imaginative and, from a mathematical point of view, unusually informal examples – of the games that economic theorists play. Thus, they provide suitable case studies for an attempted defence of model-building in economics.

I believe that each of these models tells us something important and true about the real world. My object is to discover just what these models do tell us about the world, and how they do it.

2. Akerlof and the Market for 'Lemons'

Akerlof's 1970 paper 'The market for "lemons"' is one of the best-known papers in theoretical economics. It is generally seen as having introduced to economics the concept of asymmetric information, and in doing so, sparking off what is now a whole branch of economics: the economics of information.

It is a theoretical paper that almost all economists, however untheoretical they might be, would now recognize as important. It is also a paper that just about every economic theorist would love to have written. Because there is no dispute about its value, Akerlof's paper is particularly suitable for my purposes. Everyone can see that this is a major contribution to economics.[1] The puzzle is to say exactly what the contribution is. Is Akerlof telling us anything about the real world, and if so, what?

It is worth looking closely at the structure of the paper. Here is the opening paragraph:

> This paper relates quality and uncertainty. The existence of goods of many grades poses interesting and important problems for the theory of markets. On the one hand, the interaction of quality differences and uncertainty may explain important institutions of the labour market. On the other hand, this paper presents a struggling attempt to give structure to the statement: 'business in underdeveloped countries is difficult'; in particular, a structure is given for determining the economic costs of dishonesty. Additional applications of the theory include comments on the structure of money markets, on the notion of 'insurability', on the liquidity of durables, and on brand-name goods. (Akerlof 1970: 488)

Clearly, Akerlof is claiming that his paper has something to say about an astonishingly wide range of phenomena in the real world. The paper, we are promised, is going to tell us something about the institutions of the labour market, about business in underdeveloped countries, about insurability, and so on. But what kind of thing is it going to tell us? On this point, Akerlof is rather coy. In the case of the labour market, he seems to be promising to explain some features of the real world. (Or is he? See later.) But in the case of business in underdeveloped countries, he is only going to *give structure to a statement* that is often made about the real world. Here, the implication seems to be that Akerlof's model will somehow reformulate an empirical proposition which is generally believed to be true (but might actually be false). In the other cases we are promised comments which are to be understood as applications of the theory he is to present.

Akerlof then says that, although his theory has these very general applications, he will focus on the market for used cars:

> The automobile market is used as a finger exercise to illustrate and develop these thoughts. It should be emphasized that this market is chosen for its concreteness and ease in understanding rather than for its importance or realism. (Akerlof 1970: 489)

On first reading, it is tempting to interpret 'the automobile market' as the market in which real people buy and sell real cars, and to think that Akerlof is going to present some kind of case study. One can see why he might focus on one particular market which is easy to understand, even if that market is not very important on the scale of the economy as a whole. But then what does Akerlof mean when he says that this market is not *realistic*? The object of a case study may be unrepresentative, but it cannot be unrealistic. To make sense of this passage, I think, we have to recognize that it marks a transition between the real world and the world of models. Akerlof is using the real automobile market as an example. But what he is going to present is not an empirical case study; it is a model of the automobile market. Although it is the real market which may be unimportant, it is the model which may be unrealistic.

Akerlof moves straight on to the central section of his paper, section II, entitled 'The Model with Automobiles as an Example'. The transition from reality to model is made again at the very beginning of this section:

> The example of used cars captures the essence of the problem. From time to time one hears either mention of or surprise at the large price difference between new cars and those which have just left the showroom. The usual lunch table justification for this phenomenon is the pure joy of owning a 'new' car. We offer a different explanation. Suppose (for the sake of clarity rather than realism) that there are just four kinds of cars. There are new cars and used cars. There are good cars and bad cars . . . (Akerlof 1970: 489)

The first four sentences are about an observed property of the real world: there is a large price difference between new cars and almost-new ones. Akerlof suggests that, at least from the viewpoint of the lunch table, this observation is difficult to explain. If we assume that Akerlof takes lunch with other economists, the implication is that economics cannot easily explain it; the 'pure joy' hypothesis sounds like an *ad hoc* stratagem to rescue conventional price theory. So far, then, the mode of argument might be Popperian: there is a received theory which makes certain predictions about market prices; observations of the used car market are contrary to those predictions; therefore, a new theory is needed.[2]

But from the word 'suppose' in the passage above, we move out of the real world and into the world of the model. Akerlof sets up an imaginary world; he makes no pretence to describe any real market. In this world, there are two groups of traders, 'type one' and 'type two'. All traders of a given type are alike. There are n cars, which differ only in 'quality'. Quality is measured in money units and is uniformly distributed over some range. Each group of traders maximizes an aggregate utility function. For group one, utility is the sum of the qualities of the cars it owns and the monetary value of its consumption of other goods. For group two, the utility function is the same, except that quality is multiplied by 3/2. Thus, for any given quality of car, the monetary value of a car to type one traders is less than its monetary value to type two traders. All cars are initially owned by type one traders. The quality of cars has a uniform distribution. The quality of each car is known only to its owner, but the average quality of all traded cars is known to everyone.

Akerlof admits that these assumptions are not realistic: they are not even close approximations to properties of the real used-car market. He justifies them as simplifications which allow him to focus on those features of the real market that he wishes to analyse. For example, he defends his assumptions about utility (which implicitly impose risk neutrality) against what he takes to be the more realistic alternative assumption of risk aversion by saying that he does not want to get 'needlessly mired in algebraic complication': 'The use of linear utility allows a focus on the effects of asymmetry of information; with a concave utility function we would have to deal with the usual risk-variance effects of uncertainty and the special effects we have to deal with here' (pp. 490–491).

Akerlof investigates what happens in his model world. The main conclusion is simple and startling. He shows that if cars are to be traded at all, there must be a single market price p. Then:

> However, with any price p, average quality is p/2 and therefore at no price will any trade take place at all: in spite of the fact that *at any given price* [between certain limits] there are traders of type one who are willing to sell their automobiles at a price which traders of type two are willing to pay. (Akerlof 1970: 491)

Finally, Akerlof shows what would happen in the same market if information were symmetric – that is, if neither buyers nor sellers knew the quality of individual cars, but both knew the probability distribution of quality. In this case, there is a market-clearing equilibrium price, and trade takes place, just as the standard theory of markets would lead us to expect. Akerlof ends section II at this point, so let us take stock.

What we have been shown is that in a highly unrealistic model of the used car market, no trade takes place – even though each car is worth less to its owner than it would be to a potential buyer. We have also been given some reason to think that, in generating this result, the crucial property of the model world is that sellers know more than buyers. Notice that, taken literally, Akerlof's result is too strong to fit with the phenomenon he originally promised to explain – the price difference between new and used cars.[3] Presumably, then, Akerlof sees his model as describing in extreme form the workings of some *tendency* which exists in the real used-car market, by virtue of the asymmetry of information which (he claims) is a property of that market. This tendency is a used-car version of Gresham's Law: bad cars drive out good. In the real used-car market, according to Akerlof, this tendency has the effect of reducing the average quality of cars traded, but not eliminating trade altogether; the low quality of traded cars then explains their low price.

Remarkably, Akerlof says nothing more about the *real* market in used cars. In the whole paper, the only empirical statement about the used-car market is the one I have quoted, about lunch-table conversation. Akerlof presents no evidence to support his claim that there is a large price difference between new and almost-new cars. This is perhaps understandable, since he clearly assumes that this price difference is generally known. More surprisingly, he presents no evidence that the owners of nearly-new cars know significantly more about their quality than do potential buyers. And although later in the paper he talks about market institutions which can overcome the problem of asymmetric information, he does not offer any argument, theoretical or empirical, to counter the hypothesis that such institutions exist in the used-car market. But if they do, Akerlof's explanation of price differences is undermined.

However, Akerlof has quite a lot to say about other real markets in section III of the paper, 'Examples and Applications'. In four subsections, entitled 'Insurance', 'The Employment of Minorities', 'The Costs of Dishonesty', and 'Credit Markets in Underdeveloped Countries', Akerlof presents what are effectively brief case studies. We are told that adverse selection in the insurance market is 'strictly analogous to our automobiles case' (p. 493), that 'the Lemons Principle . . . casts light on the employment of minorities' (p. 494), that 'the Lemons model can be used to make some comments on the costs of dishonesty' (p. 495), and that 'credit markets in underdeveloped countries often strongly reflect the Lemons Principle' (p. 497). These discussions are in the style that economists call 'casual empiricism'. They are suggestive, just as the used-car case is, but they cannot be regarded as any kind of test of a hypothesis. In fact, there is no hypothesis. Akerlof never defines the 'lemons principle'; all we can safely infer is that this term refers to the model of the used-car market. Ultimately, then, the claims of section III amount to this: In these four cases, we see markets that are in some way like the model.

The final part of the paper (apart from a very short conclusion) is section IV, 'Countervailing Institutions'. This is a brief discussion, again in the mode of casual empiricism, of some real-world institutions which counteract the problem of asymmetric information. The examples looked at are guarantees, brand names, hotel and restaurant chains, and certification in the labour market (such as the certification of doctors and barbers). The latter example seems to be what Akerlof was referring to in his introduction when he claimed that his approach might 'explain important institutions of the labour market'. Here, the claim seems to be that there are markets which would be like the model of the used-car market, were it not for some special institutional feature; therefore, the model explains those features.

From a Popperian perspective, sections III and IV have all the hallmarks of 'pseudo-science'. Akerlof has not proposed any hypothesis in a form that could be tested against observation. All he has presented is an empirically ill-defined 'lemons principle'. In Section III, he has assembled a fairly random assortment of evidence which appears to confirm that principle. In Section IV, he argues that the real world often is not like the model, but this is to be seen not as refutation but as additional confirmation. What kind of scientific reasoning is this?

3. Schelling's Checkerboard Model of Racial Sorting

My other example of a theoretical model in economics is not quite as famous as the market for lemons, but it is a personal favourite of mine.[4] It also deserves to be recognized as one of the earliest uses of what is now a well-established theoretical method: evolutionary game theory with localized interactions in a spatial structure. This is the chapter 'Sorting and Mixing: Race and Sex' in Schelling's book *Micromotives and Macrobehaviour* (1978).

The book as a whole is concerned with one of the classic themes of economics: the unintended social consequences of uncoordinated individual actions. Using a wide range of novel and surprising examples, Schelling sets out to show that spontaneous human interaction typically generates unintended patterns at the social level; in some cases these patterns are desirable, but in many cases they are not.

Schelling opens this chapter with an extended and informal discussion of segregation by colour and by sex in various social settings. His concern is with patterns of segregation that arise out of the voluntary choices of individuals. One important case of such self-segregation, he suggests, is the housing market of American cities. Blacks and whites[5] tend to live in separate areas; the boundaries of these areas change over time, but the segregation remains. Schelling suggests that it is unlikely that almost all Americans desire to live in such sharply segregated areas. He asks us to consider the possibility that the sharp segregation we observe at the social level is an unintended consequence of individual actions which are motivated only by a preference for not living in an area in which people of the other colour form an overwhelming majority. In the context of tables in a cafeteria for a baseball training camp, Schelling puts his hypothesis like this:

> Players can ignore, accept, or even prefer mixed tables but become uncomfortable or self-conscious, or think that others are uncomfortable or self-conscious, when the mixture is lopsided. Joining a table with blacks and whites is a casual thing, but being the seventh at a table with six players of the opposite colour imposes a threshold of self-consciousness that spoils the easy atmosphere and can lead to complete and sustained separation. (Schelling 1978: 144)

Having discussed a number of cases of self-segregation, both by colour and by sex, and in each case having floated the hypothesis that sharp segregation is an unintended consequence of much milder preferences, Schelling presents a ‘self-forming neighbourhood model’. He begins disarmingly: ‘Some vivid dynamics can be generated by any reader with a half-hour to spare, a roll of pennies and a roll of dimes, a tabletop, a large sheet of paper, a spirit of scientific enquiry, or, failing that spirit, a fondness for games’ (p. 147).

We are instructed to mark out an 8 x 8 grid of squares. The dimes and pennies:

> represent the members of two homogeneous groups – men and women, blacks and whites, French-speaking and English-speaking, officers and enlisted men, students and faculty, surfers and swimmers, the well dressed and the poorly dressed, or any other dichotomy that is exhaustive and recognizable. (Schelling 1978: 147)

We then distribute coins over the squares of the grid. Each square must either be allocated one coin or left empty (it is important to leave some empty spaces). Next, we postulate a condition which determines whether a coin is ‘content’ with its neighbourhood. For example, we might specify that a coin is content provided that at least one-third of its neighbours (that is, coins on horizontally, vertically or diagonally adjacent squares) are of the same type as itself. Then we look for coins which are not content. Whenever we find such a coin, we move it to the nearest empty square at which it *is* content (even if, in so doing, we make other coins discontented). This continues until there are no discontented coins. Schelling suggests that we try this with different initial distributions of coins and different rules. What we will find, he says, is a very strong tendency for the emergence of sharply segregated distributions of coins, even when the condition for contentedness is quite weak. I have followed Schelling’s instructions (with the help of a computer program rather than paper and coins), and I can confirm that he is right. Clearly, Schelling expects that after we have watched the workings of this model, we will find his earlier arguments about real-world segregation more convincing.

The general strategy of Schelling's chapter is remarkably similar to that of Akerlof's paper. Each author is claiming that some regularity R (bad products driving out good, persistent racial segregation with moving geographical boundaries) can be found in economic or social phenomena. Each is also claiming that R can be explained by some set of causal factors F (sellers being better-informed than buyers, a common preference not to be heavily outnumbered by neighbours not of one's own type). Implicitly, each is making three claims: that R occurs (or often occurs); that F operates (or often operates); and that F causes R (or tends to cause it). Neither presents any of these claims as a testable hypothesis, but each offers informal evidence from selected case studies which seems to support the first two claims. Each uses a formal model in support of the claim about causation. In each case, the formal model is a very simple, fully-described and self-contained world. The supposedly causal factors F are built into the specification of the model. In the model world, R is found in an extreme form. This is supposed to make more credible the claim that in the real world, F causes R. But just how is that claim made more credible?

4. Conceptual Exploration

Before going on, we need to consider an alternative reading of Akerlof and Schelling, in which their models are not intended to support any claims about the real world.[6] As Daniel Hausman (1992: 221) has pointed out, theoretical work in economics is often concerned with 'conceptual exploration' rather than 'empirical theorizing'. Conceptual exploration investigates the internal properties of models, without considering the relationship between the world of the model and the real world.

Such work can be seen as valuable, even by someone who insists that the ultimate purpose of model-building is to tell us something about the real world. For example, it can be valuable because it finds simpler formulations of existing theories, or discovers useful theorems within those theories. (Consider Paul Samuelson's demonstration that most of conventional demand theory can be deduced from a few simple axioms about consistent choice.) Or it can be valuable because it discovers previously unsuspected inconsistencies in received theories. (For example, Kenneth Arrow's impossibility theorem can be interpreted as a demonstration of the incoherence of Bergson-Samuelson welfare economics.[7]) There are also instances in which the development of a theory intended for one application has generated results which have later proved to be useful in completely different domains. (Think how much has grown out of John von Neumann and Oskar Morgenstern's exploration of strategies for playing poker.) Thus, to characterize Akerlof's and Schelling's models as conceptual exploration need not be to denigrate them.

So let us consider what we would learn from these models if we interpreted them as conceptual exploration and nothing else. Take Akerlof first. Akerlof's contribution, it might be said, is to show that some implications of the standard behavioural assumptions of economic theory are highly sensitive to the particular simplifying assumptions that are made about knowledge.[8] More specifically, the usual results about Pareto-efficient, market-clearing equilibrium trade can be radically altered if, instead of assuming that buyers and sellers are equally well-informed, we allow some degree of asymmetry of information. The message of Akerlof's paper, then, is that some commonly-invoked theoretical propositions about markets are not as robust as was previously thought. Thus, conclusions derived from models which assume symmetric information should be treated with caution, and new theories need to be developed which take account of the effects of asymmetric information. On this reading, the discussion of used cars is no more than a 'story' attached to a formal model, useful in aiding exposition and comprehension, but which can be dispensed with if necessary.[9] The paper is not about used cars: it is about the theory of markets.

What about Schelling? We might say that Schelling is presenting a critique of a commonly-held view that segregation must be the product either of deliberate public policy or of strongly segregationist preferences. The checkerboard model is a counter-example to these claims: it shows that segregation could arise without either of those factors being present. On this reading, Schelling is making an important contribution to debates about segregation in the real world, but the contribution is conceptual: he is pointing to an error in an existing theory. In terms of the symbols I introduced in section 3, Schelling is not asserting: 'R occurs, F operates, and F causes R'. All he is asserting is: 'R could occur, F could operate, and it could be the case that F caused R'.

It must be said that there is at least some textual evidence that both Akerlof and Schelling are tempted by this kind of interpretation of their models. As I have already suggested, Akerlof often seems to be taking care not to draw inferences about the real world from his model. For example, although he does claim to be offering an explanation of price differences in the real car market, his other references to 'explanation' are more nuanced. Notice that in the opening paragraph he does not claim that his model explains important institutions of the labour market: what may (not does) explain them is 'the interaction of quality differences and uncertainty'. The final sentence of the paper uses a similar formulation: 'the difficulty of distinguishing good quality from bad . . . may indeed explain many economic institutions' (p. 500). On one reading of 'may' in these passages, Akerlof is engaged only in conceptual exploration: he is considering what sorts of theory are possible, but not whether or not these theories actually explain the phenomena of the real world. However, I shall suggest that a more natural reading is that Akerlof is trying to say something like this: I believe that economists will be able to use the ideas in this paper to construct theories which *do* explain important economic institutions.

Schelling is more explicit about his method, and what it can tell us:

> What can we conclude from an exercise like this? We may at least be able to disprove a few notions that are themselves based on reasoning no more complicated than the checkerboard. Propositions beginning with 'It stands to reason that . . . ' can sometimes be discredited by exceedingly simple demonstrations that, though perhaps true, they do not exactly 'stand to reason'. We can at least persuade ourselves that certain mechanisms could work, and that observable aggregate phenomena could be compatible with types of 'molecular movement' that do not closely resemble the aggregate outcomes that they determine. (Schelling 1978: 152)

Schelling does not elaborate on what notions he has disproved. Possibly what he has in mind is the notion that either deliberate policy or the existence of strongly segregationist preferences is a necessary condition for the kind of racial segregation that is observed in American cities. His claim, then, is that he has discredited this notion by means of a counter-example.

Whatever we make of these passages, neither paper, considered as a whole, can satisfactorily be read as conceptual exploration and nothing else. The most obvious objection to this kind of interpretation is that Akerlof and Schelling both devote such a lot of space to the discussion of real-world phenomena. Granted that Akerlof's treatment of the used car market has some of the hallmarks of a theorist's 'story', what is the point of all the 'examples and applications' in his section III, or of the discussion of 'countervailing institutions' in section IV, if not to tell us something about how the world really is? This material may be casual empiricism, but it is empiricism none the less. It is not just a way of helping us to understand the internal logic of the model. Similarly, Schelling's discussion of the baseball training camp is clearly intended as a description of the real world. Its purpose, surely, is to persuade us of the credibility of the hypothesis that real people – it is hinted, people like us – have mildly segregationist preferences. If all we were being offered was a counterexample to a general theoretical claim, such material would be redundant.

Clearly, neither Akerlof nor Schelling wants to claim that his work is a completed theory. The suggestion seems to be that these are preliminary sketches of theories. The models that are presented are perhaps supposed to stand in the sort of relation to a completed theory that a 'concept car' does to a new production model, or that the clothes in a *haute couture* fashion show do to the latest designs in a fashion shop. That is, these models are suggestions about how to set about explaining some phenomenon in the real world. To put this another way, they are sketches of processes which, according to their creators, might explain phenomena we can observe in the real world. But the sense of 'might explain' here is not just the kind of logical possibility that could be discovered by conceptual exploration. (The latter sense could be paraphrased as: 'In principle, it is possible that processes with this particular formal structure could generate regularities with that particular formal structure'.) The theorist is declaring his confidence that his approach is likely to work as an explanation, even if he does not claim so to have explained anything so far.

If Akerlof's and Schelling's disclaimers were to be read as saying 'This work is conceptual exploration and nothing else', they would surely be disingenuous. We are being offered potential explanations of real-world phenomena. We are being encouraged to take these potential explanations seriously – perhaps even to do some of the work necessary to turn these sketches of theories into production models. If we are to do this, it is not enough that we have confidence in the technical feasibility of an internally consistent theory. Of course, having that confidence is important, and we can get it by conceptual exploration of formal models. But what we need in addition is some confidence that the production model is likely to do the job for which it has been designed – that it is likely to explain real-world phenomena. In other words, we need to see a sketch of an *actual* explanation, not just of a logically coherent formal structure. We should expect Akerlof's and Schelling's models to provide explanations, however tentative and imperfect, of regularities in the real world. I shall proceed on the assumption that these models are intended to function as such explanations.

5. Instrumentalism

This brings us back to the problem: How do unrealistic economic models explain real-world phenomena?

Many economists are attracted by the instrumentalist position that a theory should be judged only on its predictive power within the particular domain in which it is intended to be used. According to one version of instrumentalism, the ‘assumptions’ of a theory, properly understood, are no more than a compact notation for summarizing the theory’s predictions; thus, the question of whether assumptions are realistic or unrealistic does not arise. An alternative form of instrumentalism, perhaps more appropriate for economics, accepts that the assumptions of a theory *refer* to things in the real world, but maintains that it does not matter whether those assumptions are true or false. On either account, the assumptions of a theory *function* only as a representation of the theory’s predictions.

Instrumentalist arguments are often used in defence of the neoclassical theory of price determination which assumes utility-maximizing consumers, profit-maximizing firms, and the instantaneous adjustment of prices to market-clearing levels. In the instrumentalist interpretation the object of the neoclassical theory is to predict changes in the prices and total quantities traded of different goods as a result of exogenous changes (such as changes in technology or taxes). On this view, aggregated economic statistics play the same role in economics as the movements of the heavenly bodies through the sky did in early astronomy:[10] they are the only phenomena we want to predict, and the only (or only acceptable) data.[11] The neoclassical theory is just a compact description of a set of predictions. To ask whether its assumptions are realistic is either to make a category mistake (because assumptions do not refer to anything that has real existence) or to miss the point (because, although assumptions refer to real things, the truth or falsity of those references has no bearing on the value of the theory).

But is it possible to understand Akerlof’s and Schelling’s models instrumentally? These models are certainly similar to the neoclassical model of markets in their use of highly simplified assumptions which, if taken literally, are highly unrealistic. But if these models are intended to be read instrumentally, we should expect to find them being used to generate unambiguous predictions about the real world. Further, there should be a clear distinction between assumptions (which either have no truth values at all, or are allowed to be false) and predictions (which are asserted to be true).

In fact, neither Akerlof nor Schelling proposes any explicit and testable hypothesis about the real world. Nor does either theorist maintain an instrumentalist distinction between assumptions and predictions. Akerlof’s case studies seem to be intended as much to persuade us of the credibility of his assumptions about asymmetric information as to persuade us that the volume of trade is sub-optimal. As I have already said, Schelling’s discussion of the baseball camp seems to be intended to persuade us of the credibility of his assumptions about preferences. On the most natural readings, I suggest, Akerlof and Schelling think they are telling us about forces or tendencies which connect *real* causes (asymmetric information, mildly segregationist preferences) to *real* effects (sub-optimal volumes of trade, sharp segregation). Akerlof's and Schelling's unrealistic models are supposed to give support to these claims about real tendencies. Whatever method this is, it is not instrumentalism: it is some form of realism.

6. Metaphor and Caricature

Allan Gibbard and Hal Varian (1978) offer an interpretation of economic models which emphasizes explanation rather than prediction. They characterize a model as the conjunction of two elements: an uninterpreted formal system within which logical deductions can be made, and a 'story' which gives some kind of interpretation of that formal system. With Schelling's checkerboard model apparently in mind, they describe a form of modelling in which the fit of the model to the real world is *casual*:

> The goal of casual application is to explain aspects of the world that can be noticed or conjectured without explicit techniques of measurement. In some cases, an aspect of the world (such as price dispersal, housing segregation, and the like) is noticed, and certain aspects of the micro-situation are thought perhaps to explain it; a model is then constructed to provide the explanation. In other cases, an aspect of the micro-world is noticed, and a model is used to investigate the kinds of effects such a factor could be expected to have. (Gibbard and Hal Varian 1978: 672)

This seems a fair description of what both Akerlof and Schelling are doing. But Gibbard and Varian have disappointingly little to say about *how* a casual model explains an aspect of the real world, or how it allows us to investigate the likely effects of real-world factors on real-world phenomena.

Gibbard and Varian recognize – indeed, they welcome – the fact that casual models are unrealistic; but their defence of this lack of realism is itself rather casual:

> When economic models are used in this way to explain casually observable features of the world, it is important that one be able to grasp the explanation. Simplicity, then, will be a highly desirable feature of such models. Complications to get as close as possible a fit to reality will be undesirable if they make the model less possible to grasp. Such complications may, moreover, be unnecessary, since the aspects of the world the model is used to explain are not precisely measured. (Gibbard and Hal Varian 1978: 672)

The suggestion here seems to be that the purpose of a model is to communicate an idea to an audience; simplicity is a virtue because it makes communication easier. But this puts the cart before the horse. What has to be communicated is not just an idea: it is a claim about how things really are, along with reasons for accepting that claim as true. Simplicity in communication has a point only if there is something to be communicated. While granting that Akerlof's and Schelling's models are easy to grasp, we may still ask what exactly we have grasped. How do these models come to be explanations? And explanations of what?

One possible answer is given by Deirdre McCloskey (1983: 502–507), who argues that models are metaphors. According to McCloskey, the modeller's claim is simply that the real world is like the model in some significant respect (p. 502). In evaluating a model, we should ask the same questions as we would when evaluating a metaphor: 'Is it illuminating, is it satisfying, is it apt?' (p. 506). The claim 'models are metaphors' must, I think, be understood as a metaphor in itself. As a metaphor, it is certainly satisfying and apt; but, in relation to our examination of Akerlof's and Schelling's models, just how illuminating is it?

Clearly, Akerlof and Schelling are claiming that the real world is like their models in some significant respects. What is at issue is what exactly these claims amount to, and how (if at all) they can be justified. Translating into McCloskey's language, what is at issue is how illuminating and how apt Akerlof's and Schelling's metaphors are. But this translation of the question does not take us any nearer to an answer.

Gibbard and Varian (1978) come closer to giving an answer to this question (at this stage, I do not say the right answer) when they suggest that models are *caricatures*. The concept of caricature is tighter than that of metaphor, since the ingredients of a caricature must be taken from the corresponding reality. (Compare cartoons – John Bull, the fat, beef-eating yeoman farmer, was originally a caricature of a characteristic Englishman. Although no longer a valid caricature, he is still recognizable as a symbol of, or metaphor for, Englishness.) According to Gibbard and Varian, the assumptions of a model may be chosen 'not to approximate reality, but to exaggerate or isolate some feature of reality' (p. 673). The aim is 'to distort reality in a way that illuminates certain aspects of that reality' (p. 676).

The idea that models are caricatures suggests that models may be able to explain the real world because their assumptions describe certain features of that world, albeit in isolated or exaggerated form. Gibbard and Varian do not pursue this idea very far, but it is taken up in different ways by Hausman (1992: 123–151) and by Uskali Mäki (1992, 1994), whose work will now be discussed.

7. Economics as an Inexact Deductive Science, and the Method of Isolation

I have suggested that Akerlof and Schelling are each pointing to some tendency in the real world, which each claims to explain by means of a model. One way of trying to make sense of the idea of 'tendencies' is by means of what Hausman calls 'implicit *ceteris paribus* clauses'. The underlying idea is that the phenomena of the real world are the product of the interaction of many different causal factors. A tendency (some writers prefer the term 'capacity') is to be understood as the workings of some small subset of these factors.

In order to describe a tendency, we must somehow isolate the relevant subset of factors from the rest. Thus, the description is expressed in counterfactual terms, such as 'in the absence of all other causal factors, L' or 'if all other causal factors are held constant, L' where L is some law-like proposition about the world. Hausman argues that in economics, *ceteris paribus* clauses are usually both implicit and vague. He uses the term inexact generalization for generalizations that are qualified by implicit *ceteris paribus* clauses.

Hausman argues that economics arrives at its generalizations by what he calls the inexact deductive method. He summarizes this method as the following four-step schema:

1. *Formulate* credible (*ceteris paribus*) and pragmatically convenient generalizations concerning the operation of relevant causal variables;
2. *Deduce* from these generalizations, and statements of initial conditions, simplifications, etc., predictions concerning relevant phenomena;
3. *Test* the predictions;
4. If the predictions are correct, then regard the whole amalgam as confirmed. If the predictions are not correct, then *compare* alternative accounts of the failure on the basis of explanatory success, empirical progress, and pragmatic usefulness (p. 222).

For Hausman, this schema is 'both justifiable and consistent with existing theoretical practice in economics, insofar as that practice aims to appraise theories empirically' (p. 221).[12] By following this schema, economists can arrive at inexact generalizations about the world, which they are entitled to regard as confirmed. The schema is an adaptation of John Stuart Mill's (1843, Book 6, chs 1–4) account of the 'logic of the moral sciences'. (The most significant amendment is that, in Hausman's schema, the premises from which deductions are made are merely 'credible generalizations' which may be called into question if the predictions derived from them prove false. In contrast, Mill seems to have thought that the inexact predictions of economics could be deduced from proven 'laws of mind'.)

Mäki's account of how economic theories explain reality has many similarities with Hausman's. Like Hausman, Mäki argues that theoretical assumptions should be read as claims about what is true in the real world. But where Hausman talks of *inexact* propositions, Mäki talks of *isolations.* Economics, according to Mäki, uses 'the method of isolation, whereby a set of elements is theoretically removed from the influence of other elements in a given situation' (1992: 318). On this account, a theory represents just some of the factors which are at work in the real world; the potential influence of other factors is 'sealed off' (p. 321). Such sealing-off makes a theory unrealistic; but the theory may still claim to describe an aspect of reality.

As Mäki (p. 325) notices, there is a parallel between his concept of theoretical isolation and the idea of *experimental* isolation. Laboratory experiments investigate particular elements of the world by isolating them; the mechanisms by which other elements are sealed off are experimental controls. The laboratory environment is thereby made unrealistic, in the sense that it is 'cleaner' than the world outside; but this unrealisticness is an essential feature of the experimental method. On this analogy, models are *thought experiments.*[13]

But if a thought experiment is to tell us anything about the real world (rather than merely about the structure of our own thoughts), our reasoning must in some way replicate the workings of the world. For example, think how a structural engineer might use a theoretical model to test the strength of a new design. This kind of modelling is possible in engineering because the theory which describes the general properties of the relevant class of structures is already known, even though its implications for the new structure are not. Provided the predictions of the general theory are true, the engineer's thought experiment replicates a physical experiment that could have been carried out.

On this interpretation, then, a model explains reality by virtue of the truth of the assumptions that it makes about the causal factors it has isolated. The isolations themselves may be unrealistic; in a literal sense, the assumptions which represent these isolations may be (and typically are) false. But the assumptions which represent the workings of the isolated causal factors need to be true. So, I suggest, the implications of the method of isolation for theoretical modelling are broadly similar to the first two steps of Hausman's schema. That is, the modeller has to formulate credible generalizations concerning the operation of the factors that have been isolated, and then use deductive reasoning to work out what effects these factors will have in particular controlled environments.

So is this what Akerlof and Schelling are doing? Even though neither author explicitly proposes a testable hypothesis, we might perhaps interpret them as implicitly proposing *ceteris paribus* hypotheses. (Later, I shall suggest what these hypotheses might be.) But if Akerlof's and Schelling's models are to be understood as instances of the inexact deductive method, each model must be interpreted as the deductive machinery which generates the relevant hypothesis. For such an interpretation to be possible, we must be able to identify the simplifying assumptions of the model with the *ceteris paribus* or non-interference clauses of the hypothesis. That is, if the hypothesis takes the form 'X is the case, provided there is no interference of types $i_1, \ldots, i_n$', then the model must deduce X from the conjunction of two sets of assumptions. The first set contains 'credible and pragmatically convenient generalizations' – preferably ones which have been used successfully in previous applications of the inexact deductive method. The second set of assumptions – which Mäki would call 'isolations' – postulate the non-existence of $i_1, \ldots, i_n$.

Take Akerlof's model. Can its assumptions be understood in this way? Some certainly can. For example, Akerlof implicity assumes that each trader maximizes expected utility. Correctly or incorrectly, most economists regard expected utility maximization as a well-grounded generalization about human behaviour; there are (it is thought) occasional exceptions, but these can safely be handled by implicit non-interference clauses. Similarly, Akerlof assumes that if an equilibrium price exists in a market, that price will come about, and the market will clear. This, too, is a generalization that most economists regard as well-grounded. There is a standing presumption in economics that, if an empirical statement is deduced from standard assumptions such as expected utility maximization and market-clearing, then that statement is reliable: the theorist does not have to justify those assumptions anew in every publication.

As an example of the other type of assumption, notice that Akerlof's model excludes all of the 'countervailing institutions' which he discusses in his section IV. Presumably, if Akerlof is proposing an empirical hypothesis, it must be something like the following: 'If sellers know more than buyers about the quality of a good, and if there are no countervailing institutions, then the average quality of those goods that are traded is lower than that of goods in general.' The absence of countervailing institutions is a non-interference clause in the hypothesis, and therefore also a legitimate property of the model from which the hypothesis is deduced.

The difficulty for a Hausman-like or Mäki-like interpretation is that Akerlof's and Schelling's models both include many assumptions which neither are well-founded generalizations nor correspond with *ceteris paribus* or non-interference clauses in the empirical hypothesis that the modeller is advancing. Akerlof assumes that there are only two types of trader, that all traders are risk-neutral, that all cars are alike except for a one-dimensional index of quality, and so on. Schelling assumes that all individuals are identical except for colour, that they live in the squares of a rectangular grid, and so on again. These are certainly not well-founded empirical generalizations. So can they be read as *ceteris paribus* clauses?

If we are to interpret these assumptions as ceteris paribus clauses, there must be corresponding restrictive clauses in the hypotheses that are deduced from the models. That is, we must interpret Akerlof and Schelling as proposing counterfactual empirical hypotheses about what would be observed, were those assumptions true. But if we pursue the logic of this approach, we end up removing almost all empirical content from the implications of the models – and thereby defeating the supposed objective of the inexact deductive method. Take the case of Schelling's model. Suppose we read Schelling as claiming that *if* people lived in checkerboard cities, and *if* people came in just two colours, and *if* each person was content provided that at least a third of his neighbours were the same colour as him, and *if*..., and *if*... (going on to list all the properties of the model), *then* cities would be racially segregated. That is not an empirical claim at all: it is a theorem.

Perhaps the best way to fit Akerlof's and Schelling's models into Hausman's schema is to interpret their troublesome assumptions as the 'simplifications etc.' referred to in step 2 of that schema. But this just shunts the problem on, since we may then ask why it is legitimate to introduce such simplifications into a deductive argument. The conclusions of a deductive argument cannot be any stronger than its premises. Thus, any hypothesis that is generated by a deductive method must have implicit qualifying clauses corresponding with the assumptions that are used as premises. And this does not seem to be true of Akerlof's and Schelling's hypotheses.

To understand what Akerlof and Schelling are doing, we have to realize that results that they derive deductively within their models are not the same as the hypotheses that they want us to entertain. Consider exactly what Akerlof and Schelling are able to show by means of their models. Akerlof shows us that under certain specific conditions (there are just two types of trader, all cars are identical except for quality, sellers' valuations of cars of given quality are two-thirds those of buyers, etc.), no trade takes place. Among these conditions is a particular assumption about asymmetric information: sellers know the quality of their cars, but buyers don't. Akerlof also shows that if the only change that is made to this set of conditions is to assume symmetric information instead of asymmetric, then trade does take place. Thus, Akerlof has proved a *ceteris paribus* result, but only for a particular array of other conditions. This result might be roughly translated as the following statement: If all other variables are held constant at the particular values assumed in the model, then an increase in the degree of asymmetry of information reduces the volume of trade.

What about Schelling? Schelling shows – or, strictly speaking, he invites us to show ourselves – that under certain specific conditions (people come in just two colours, each person is located on a checkerboard, etc.) individuals' independent choices of location generate segregated neighbourhoods. Among these conditions is a particular assumption about individuals' preferences concerning the colour composition of their neighbourhoods: people prefer not to live where more than some proportion p of their neighbours are of the other colour. Schelling invites us to try out different values of p. We find that segregated neighbourhoods eventually evolve, whatever value of p we use, provided it is less than 1. If $p = 1$, that is, if people are completely indifferent about the colours of their neighbours, then segregated neighbourhoods will not evolve. (Schelling does not spell out this latter result, but a moment's thought about the model is enough to derive it.) Thus, we have established a *ceteris paribus* result analogous with Akerlof's: we have discovered the effects of changes in the value of p, when all other variables are held constant at the particular values specified by the model.

To put this more abstractly, let x be some variable whose value we are trying to explain, and let $(v_1, \ldots, v_n)$ be an array of variables which might have some influence on x. What Akerlof and Schelling each succeed in establishing by deductive reasoning is the truth of a proposition of the form: If the values of $v_2, \ldots, v_n$ are held constant at the specific values $v_2^*, \ldots, v_n^*$, then the relationship between v_1 and x is.... The values $v_2^*, \ldots, v_n^*$ are those built into the relevant model. Taken at face value, this proposition tells us nothing about the relationship between v_1 and x in the actual world. It tells us only about that relationship in a counterfactual world.

But Akerlof and Schelling want us to conclude that certain much more general propositions are, if not definitely true, at least credible. When Akerlof talks about the 'lemons principle', he has in mind some broad generalization, perhaps something like the following: For all markets, if all other features are held constant, an increase in the degree of asymmetry of information reduces the volume of trade. Similarly, what Schelling has in mind is some generalization like the following: For all multi-ethnic cities, if people prefer not to live in neighbourhoods where the vast majority of their neighbours are of another ethnic group, strongly segregated neighbourhoods will evolve. In my more abstract notation, the generalizations that Akerlof and Schelling have in mind have the form: If the values of $v_2, \ldots, v_n$ are held constant at any given value, then the relationship between v_1 and x is. . . .

If these generalizations are to be interpreted as hypotheses, the models are supposed to give us reasons for thinking that they are true. If the generalizations are to be interpreted as observed regularities, the models are supposed to explain why they are true. But deductive reasoning cannot fill the gap between the specific propositions that can be shown to be true in the model world (that is, propositions that are true if $v_2, \ldots, v_n$ are held constant at the values $v_2^* \ldots, v_n^*$) and the general propositions that we are being invited to entertain (that is, those that are true if $v_2, \ldots, v_n$ are held constant at any values). Somehow, a transition has to be made from a particular hypothesis, which has been shown to be true in the model world, to a general hypothesis, which we can expect to be true in the real world too.

8. Inductive Inference

So how can this transition be made? As before, let R stand for a regularity (bad products driving out good, persistent racial segregation with moving geographical boundaries) which may or may not occur in the real world. Let F stand for a set of causal factors (sellers being better-informed than buyers, a common preference not to be heavily outnumbered by neighbours not of one's own type) which may or may not operate in the real world. Akerlof and Schelling seem to be reasoning something like this:

Schema 1: Explanation
E1 – in the model world, R is caused by F.
E2 – F operates in the real world.
E3 – R occurs in the real world.
Therefore, there is reason to believe:
E4 – in the real world, R is caused by F.

Alternatively, if we read Akerlof and Schelling as implicitly proposing empirical hypotheses, we might represent their reasoning as:

Schema 2: Prediction
P1 – in the model world, R is caused by F.
P2 – F operates in the real world.
Therefore, there is reason to believe:
P3 – R occurs in the real world.

A third possible reading of Akerlof and Schelling involves abductive reasoning (inferring causes from effects):[14]

Schema 3: Abduction
A1 – in the model world, R is caused by F.
A2 – R occurs in the real world.
Therefore, there is reason to believe:
A3 – F operates in the real world.

In each of these three reasoning schemata, the ‘therefore’ requires an inductive leap. By ‘induction’ I mean any mode of reasoning which takes us from specific propositions to more general ones (compare the similar definition given by Mill [1843, Book 3, ch. 1, p. 186]). Here, the specific proposition is that R is caused by F in the case of the model. In order to justify each of the ‘therefores’, we must be justified in inferring that R is caused by F more generally. *If* there is a general causal link running from F to R, then when we observe F and R together in some particular case (that is, the case of the real world), we have some reason to think that the particular R is caused by the particular F (explanation). Similarly, when we observe F in a particular case, we have some reason to expect to find R too (prediction). And when we observe R in a particular case, we have some reason to expect to find F too (abduction). It seems, then, that Akerlof’s and Schelling’s method is not purely deductive: it depends on induction as well as on deduction. But how might these inductions be justified?

9. Justifying Induction: Separability

One possible answer is to appeal to a very general hypothesis about causation, which (to my knowledge) was first invoked by Mill (1843, Book 3, ch. 6, pp. 242–247). Mill defines phenomena as mechanical if the overall effect of all causal factors can be represented as an addition of those separate factors, on the analogy of the vector addition of forces in Newtonian physics. Given this hypothesis of the composition of causes, we are entitled to move from the *ceteris paribus* propositions which have been shown to be true in a model to more general *ceteris paribus* propositions which apply to the real world too.[15] Using the notation introduced in section 6, this immediately closes the gap between a proposition which is true if certain variables $v_2, \ldots, v_n$ are held constant at certain specific values $v_2^*, \ldots, v_n^*$ and a proposition which is true if $v_2, \ldots, v_n$ are held constant at any values: if the proposition is true in the first case, then (if the hypothesis about the composition of causes is true) it is true in the second case too. But what entitles us to use that hypothesis itself?

In some cases, it may be legitimate to treat that hypothesis as a proven scientific law – as in the paradigm case of the composition of forces in physics. Mill seems to have taken it to be an *a priori* truth that 'In social phenomena the Composition of Causes is the universal law' (1843, Book 6, ch. 7, p. 573). However, the argument Mill gives in support of this claim is quite inadequate. He simply asserts that 'Human beings in society have no properties but those which are derived from, and may be resolved into, the laws of the nature of individual man'. But even if we grant this assertion, all we have established is that social facts are separable into facts about individuals. We have not established the separability of *causal factors*. Thus, for example, the fact that society is an aggregate of individuals does not allow us to deduce that if an increase in the price of some good in one set of circumstances causes a decrease in consumption, then the same cause will produce the same effect in other circumstances.

Hausman (1992: 138) offers a defence for Mill's method in economics. He claims that Mill's supposition that economic phenomena are mechanical is 'implicit in most applications of economic models', and then says: 'Its only justification is success'. In other words, this supposition is an inductive inference from the general experience of economic modelling.

But this argument seems to beg the question. For the sake of the argument, let us grant that economic modelling has often been successful – successful, that is, in relation to Hausman's criterion of generating correct predictions about the real world. Even so, the explanation of its success may be that economists are careful not to rely on models unless they have some independent grounds for believing that the *particular* phenomena they are trying to explain are mechanical – or, more generally, unless they have some independent grounds for making particular inductive inferences from the world of the model to the real world. Given the *prima facie* implausibility of the assumption that all economic phenomena are mechanical, it would be surprising to find that this assumption was the main foundation for inductive inferences from theoretical models. We should look for other foundations.

10. Justifying Induction: Robustness

One way in which inductions might be justified is by showing that the results derived from a model are *robust* to changes in the specification of that model. Gibbard and Varian (1978: 675) appeal to the robustness criterion when they suggest that, in order for caricature-like models to help us to understand reality, 'the conclusions [should be] robust under changes in the caricature'. Hausman (1992: 149) makes a somewhat similar appeal when he considers the conditions under which it is legitimate to use simplifications – that is, propositions that are not true of the real world – in the second stage of his schema of the inexact deductive method. He proposes a set of conditions which he glosses as 'reasonable criteria for judging whether the falsity in simplifications is irrelevant to the conclusions one derives with their help'.

One significant implication of this approach is that simplifications need not be isolations. Take Schelling's checkerboard city. The simplicity of the checkerboard city lies in the way that its pattern repeats itself: if we ignore the edges of the board, every location is identical with every other. (More showy theorists than Schelling would probably draw the checkerboard on a torus, so that it had no edges at all; this would give us a city located on a doughnut-shaped planet.) This property of 'repeatingness' makes the analysis of the model much easier than it otherwise would be. But it does not seem right to say that the checkerboard *isolates* some aspect of real cities by sealing off various other factors which operate in reality: just what do we have to seal off to make a real city – say, Norwich – become like a checkerboard? Notice that, in order to arrive at the checkerboard plan, it is not enough just to suppose that all locations are identical with one another (that is, to use a 'generic' concept of location): we need to use a *particular form* of generic location. So, I suggest, it is more natural to say that the checkerboard plan is something that Schelling has *constructed* for himself. If we think that Schelling's results are sufficiently robust to changes in the checkerboard assumption, that assumption may be justified, even though it is not an isolation.[16]

Robustness arguments work by giving reasons for believing that a result that has been derived in one specific model would also be derived from a wide class of models, or from some very general model which included the original model as a special case. Economic theorists tend to like general models, and much effort is put into generalizing results. By experience, theorists pick up a feel for the kinds of result that can be generalized and the kinds that cannot be. The main way of making this distinction, I think, is to examine the links between the assumptions of a model and its results, and to try to find out which assumptions are (as theorists say) 'doing the work'. If a model has already been presented in a somewhat general way, it is often useful to strip it down to its simplest form, and then to see which assumptions are most closely associated with the derivation of the relevant result.[17]

In both Akerlof's and Schelling's models, there are good reasons to think that most of the simplifying assumptions are orthogonal to the dimension on which the model 'works': these are simplifying assumptions which could be changed or generalized without affecting the qualitative results. In many cases, Akerlof argues exactly this. Recall, for example, his discussion of risk neutrality. Akerlof could have assumed risk aversion instead, which would have made the model much less easy to work with; but there does not seem to be any way in which the major qualitative conclusions are being driven by the assumption of risk neutrality. Similarly, in the case of Schelling's model, the checkerboard layout seems to have nothing particular to do with the tendency for segregation. Schelling is confident enough to invite the reader to try different shapes of boards, and might easily have suggested different tessellations (such as triangles or hexagons).

Notice how this mode of reasoning remains in the world of models – which may help to explain why theorists feel comfortable with it. It makes inductive inferences from one or a small number of models to *models* in general. For example: having experimented with Schelling's checkerboard model with various parameter values, I have found that the regularity described by Schelling persistently occurs. Having read Schelling and having thought about these results, I think I have some feel for why this regularity occurs; but I cannot give any proof that it *must* occur (or even that it must occur with high probability). My confidence that I would find similar results were I to use different parameter values is an inductive inference. I also feel confident (although not quite as confident as in the previous case) that I would find similar results if I used triangles or hexagons instead of squares. This is an inductive inference too.

Obviously, however, it cannot be enough to stay in the world of models. If the theorist is to make claims about the real world, there has to be some link between those two worlds. For example, it is not enough to be convinced that what Schelling has shown us to be true of checkerboard cities is also true of other modelcities: we have to be convinced that it is true of real cities. We have to think something like the following: If what Schelling has shown us is true of checkerboard cities, then it will probably tend to be true of cities in general. What makes that inductive inference credible?

11. Justifying Induction: Credible Worlds

Inductive reasoning works by finding some regularity R in some specific collection of observations $x_1, \ldots, x_n$, and then inferring that the same regularity will probably be found throughout a general set of phenomena S, which contains not only $x_1, \ldots, x_n$ but also other elements which have not yet been observed. For example, $x_1, \ldots, x_n$ might be the n different versions of Schelling's checkerboard city that I have so far experimented with, R might be the emergence of segregation in model cities, and S might be the set of all checkerboard cities. Having found R in the n particular cities, I infer that this is a property of checkerboard cities in general.

Unavoidably, inductive reasoning depends on prior concepts of similarity: we have to be able to interpret S as the definition of some *relevant* or *salient* respect in which $x_1, \ldots, x_n$ are similar. Many of the philosophical puzzles surrounding induction stem from the difficulty of justifying any criterion of similarity.[18] Obviously, I am not going to solve these deep puzzles towards the end of a paper about models in economics.[19] For my purposes, what is important is this: if we are to make inductive inferences from the world of a model to the real world, we must recognize some significant similarity between those two worlds.

If we interpret Akerlof and Schelling as using schema 1 or schema 2 (see section 7), it might be said that this similarity is simply the set of causal factors F: what the two worlds have in common is that those factors are present in both. To put this another way, the real world is equivalent to an immensely complicated model: it is the limiting case of the process of replacing the simplifying assumptions of the original model with increasingly realistic specifications. If (as I argued in section 10) we can legitimately make inductive inferences from a simple model to slightly more complex variants, then we must also have *some* warrant for making inferences to much more complex variants, and hence also to the real world. Nevertheless, the enormous difference in complexity between the real world and any model we can hope to analyse – and hence the apparent lack of similarity between the two – suggests that we ought to be very cautious about making inferences from the latter to the former.

So what might increase our confidence in such inferences? I want to suggest that we can have more confidence in them, the greater the extent to which we can understand the relevant model as a description of how the world *could be.*

Let me explain. Inductive inferences are most commonly used to take us from one part of the real world to another. For example, suppose we observe racial segregation in the housing markets of Baltimore, Philadelphia, New York, Detroit, Toledo, Buffalo and Pittsburgh. Then we might make the inductive inference that segregation is a characteristic of large industrial cities of the north-eastern USA, and so form the expectation that there will be segregation in say, Cleveland. Presumably, the thought behind this inference is that the forces at work in the Cleveland housing market, whatever these may be, are likely to be broadly similar to those at work in other large industrial cities in north east USA. Thus, a property that is true for those cities in general is likely to be true for Cleveland in particular. One way of describing this inference is to say that each of the housing markets of Baltimore, Philadelphia, New York, etc. constitutes a *model* of the forces at work in large industrial north-eastern US cities. These, of course, are natural models, as contrasted with *theoretical* models created in the minds of social scientists. But if we can make inductive inferences from natural models, why not from theoretical ones? Is the geography of Cleveland any more like the geography of Baltimore or Philadelphia than it is like the geography of Schelling's checkerboard city?[20]

What Schelling has done is to construct a set of *imaginary* cities, whose workings 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 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 those preferences.

Compare Akerlof. Akerlof has constructed two variants of an imaginary used-car market. In one variant, buyers and sellers have the same imperfect information about the quality of cars, and trade takes place quite normally. In the other variant, sellers know more than buyers, and no trade takes place at all. When we think about how these markets work, it becomes credible to suppose that many variant imaginary markets can be constructed, and that these share the common feature that, *ceteris paribus*, the volume of trade falls as information becomes less symmetric. We are invited to make the inductive inference that similar causal processes apply in real markets, with similar effects. Thus in real markets too, *ceteris paribus*, the volume of trade is positively related to the symmetry of information.

We gain confidence in such inductive inferences, I suggest, by being able to see the relevant models as instances of some category, some of whose instances actually exist in the real world. Thus, we see Schelling's checkerboard cities as *possible cities*, alongside real cities like New York and Philadelphia. We see Akerlof's used-car market as a *possible market*, alongside real markets such as the real market for used cars in a particular city, or the market for a particular type of insurance. We recognize the significance of the similarity between model cities and real cities, or between model markets and real markets, by accepting that the model world *could be* real – that it describes a state of affairs that is *credible*, given what we know (or think we know) about the general laws governing events in the real world. On this view, the model is not so much an abstraction from reality as a parallel reality. The model world is not constructed by starting with the real world and stripping out complicating factors: although the model world is simpler than the real world, the one is not a *simplification* of the other.

Credibility in models is, I think, rather like credibility in 'realistic' novels. In a realistic novel, the characters and locations are imaginary, but the author has to convince us that they are credible – that there could be people and places like those in the novel. As events occur in the novel, we should have the sense that these are natural outcomes of the way the characters think and behave, and of the way the world works. We judge the author to have failed if we find a person acting out of character, or if we find an anachronism in a historical novel: these are things that *couldn't* have happened. But we do not demand that the events of the novel did happen, or even that they are simplified representations of what really happened. (Simplification and isolation are allowed, of course; we do not expect to be told everything that the characters do or think. But what is being simplified is not the world of actual events, but the world imagined by the author.) We can praise a novel for being 'true to life' while accepting that every event within it is fictional, as when we recognize aspects of its characters as typical of people we know. When a novel has this form of truth, we can even use it to explore 'What would happen if . . . ?' questions, in something like the same way that economists can use models. By following the characters' reactions to events that we have not ourselves experienced, we may gain insights into how we would react in similar circumstances.[21]

But the reader will expect more than analogy. The obvious question that I have to answer is: What constitutes credibility in economic models? I cannot give anything remotely like a complete answer; the best I can offer are a few criteria that have guided me in my own work as a modeller, and which are exemplified in the economic models that I most admire.

For me, one important dimension of credibility is *coherence*. Everyone recognizes that a theoretical model has to be *logically* coherent, but I mean something more than this. The assumptions of a good model cohere in the broader sense that they fit naturally together. For example, some economic models assume that agents are well-informed and highly rational, while others assume that agents are poorly-informed and follow rough rules of thumb. Which type of model is more useful in explaining particular phenomena is a matter of judgement. But a model which uses an apparently arbitrary mix of the two kinds of assumption – assuming hyper-rationality in one context and bounded rationality in another – has the same kind of fault as a novel in which someone acts out of character. If a model lacks coherence, its results cannot be seen to follow naturally from a clear conception of how the world might be; this prompts the suspicion that the assumptions have been cobbled together to generate predetermined results. *Ad hoc* models of this kind may be commonplace in economics journals, but if they are, that does not justify them.

For a model to have credibility, it is not enough that its assumptions cohere with one another; they must also cohere with what is known about causal processes in the real world. Thus, Akerlof's assumption that prices tend to their market-clearing levels is justified by evidence from a wide range of 'natural' and laboratory markets. Schelling's assumption that many people have at least mildly segregationist preferences is justified by psychological and sociological evidence, and coheres with common intuition and experience. However, it is not necessary that the assumptions of the model correspond with – or even with a simplification of – any *particular* real-world situation. Thus, we should not object to Akerlof's assumption that traders' utility functions are additively separable in money and the quality of cars, or his assumption that cars are worth exactly 50 per cent more to traders of one type than they are to traders of another. These are *restrictive* assumptions, but they seem adequately *representative* of people who trade cars in the real world. In the same way, the author of a novel might choose to call her principal character Frank, make him 48 years old, and fix his home town as Ipswich. If the logic of the novel requires only that the principal character is middle-aged, male and English, there is a sense in which this specification is highly restrictive; but the character has to have *some* name, *some* age, and *some* home town, and this particular specification is adequately representative of middle-aged English men (whereas, say, naming the character Duck Bill Platypus is not).

Akerlof in particular puts a lot of effort into making his model credible in the sense I have tried to describe. The world of his model is much more uniform and regular than the real world, but Akerlof clearly wants us to think that there *could* be a used-car market which was like his model. The 'cars' and 'traders' of his model are not just primitives in a formal deductive system. They are, I suggest, cars which are *like* real cars, and traders which are *like* real traders, inhabiting a world which Akerlof has imagined, but which is sufficiently close to the real world that we can imagine its being real. Recall the sentence in which Akerlof seems to slip between talking about the real used-car market and talking about his model: the fact that such slippage is possible may be an indication that Akerlof has come to think of his model as if it were real.

At first sight, Schelling seems rather less concerned to make us believe in his model world as a possible reality. Instead of following Akerlof's strategy of basing his model on one typical case, Schelling almost always refers to the two types of actor in his model as 'dimes' and 'pennies'. But this is perhaps dictated by Schelling's strategy of asking the reader to perform the actions in the model: he has to say 'now move that dime' rather than 'that dime now moves'. Possibly, too, it reflects an embarrassment about dealing directly with the issue of racial prejudice. But when Schelling describes the laws of motion of these coins, it is clear that we are expected to think of them as people. For example, one of his suggestions is that 'we can postulate that every dime wants at least half its neighbours to be dimes, every penny wants a third of its neighbours to be pennies, and any dime or penny whose immediate neighbourhood does not meet these conditions gets up and moves' (pp. 147–148). Or again, officially referring to a dime or penny in a world of dimes and pennies: 'He is content or discontent with his neighbourhood according to the colours of the occupants of those eight surrounding squares . . . ' (p. 148). Even allowing for the fact that the use of 'he' and 'colour' rather than 'it' and 'type of coin' are probably slips, it is surely obvious that Schelling wants us to think of the dimes and pennies as people of two groups who have some embarrassment about being together. Similarly, we are expected to think of the checkerboard as a city (or some other social space, such as a dining room). Further, we are encouraged to think of these people's attitudes to one another as credible and understandable – even forgivable (recall the passage about mixed tables in the cafeteria, which precedes the checkerboard model). What Schelling has constructed is a model city, inhabited by people who are *like* real people.

12. Conclusion

I have referred several times to a puzzling common feature of the two papers. Both authors seem to want to make empirical claims about properties of the real world, and to want to argue that these claims are supported by their models. But on closer inspection of the texts, it is difficult to find any explicit connection being made between the models and the real world. Although both authors discuss real-world phenomena, neither seems prepared to endorse any specific inference from his model, still less to propose an explicit hypothesis which could be tested.

I suggest that the explanation of this puzzle is that Akerlof and Schelling are engaged in a kind of theorizing the usefulness of which depends on inductive inferences from the world of models to the real world. Everyone makes inductive inferences, but no one has really succeeded in justifying them. Thus, it should not be surprising if economists leave gaps in their explicit reasoning at those places where inductive inferences are required, and rely on their readers using their own intuitions to cross those gaps. Nor should it be surprising if economists use rhetorical devices which tend to hide these gaps from view.

Nevertheless, the gap between model and real world has to be bridged. If a model is genuinely to tell us something, however limited, about the real world, it cannot be *just* a description of a self-contained imaginary world. And yet theoretical models in economics often *are* descriptions of self-contained and imaginary worlds. These worlds have not been formed merely by abstracting key features from the real world; in important respects, they have been *constructed* by their authors.

The suggestion of this paper is that the gap between model world and real world can be filled by inductive inference. On this account, models are not internally consistent sets of uninterpreted theorems; but neither are they simplified or abstracted or exaggerated descriptions of the real world. They describe credible counterfactual worlds. This credibility gives us some warrant for making inductive inferences from model to real world.

Acknowledgements

A previous version of this paper was prepared for the conference *Fact or Fiction? Perspectives on Realism and Economics* at the Erasmus Institute for Philosophy and Economics, Rotterdam, in November 1997. The paper has been much improved as a result of the discussion at that conference. I particularly thank Nancy Cartwright, Stephan Hartmann, Daniel Hausman, Maarten Janssen, Uskali Mäki, Mary Morgan and Chris Starmer for advice. I did most of the work on the paper while visiting the Centre for Applied Ethics at the University of British Columbia, for whose hospitality I am grateful. Subsequent work was supported by the Leverhulme Trust.

Notes

1. But it was not immediately recognized as a major contribution: it was turned down three times before being accepted for publication. Mark Blaug (1997) uses this fact to suggest that Akerlof's paper is the exception which proves the rule – the rule being that modern economics is becoming 'an intellectual game played for its own sake and not for its practical consequences', creating models which are 'scandalously unrepresentative of any recognizable economic system' (pp. 2–4). However, he does not explain why Akerlof is to be acquitted of this charge.
2. An alternative reading is possible. Akerlof never claims outright that the 'pure 'joy' explanation is false, or that his own explanation is correct – only that it is 'different'. So could it be that he doesn't want to make any such claims? In section 3, I consider – and reject – the suggestion that Akerlof is not claiming to explain any features of the real world.
3. Akerlof deals with this problem to some degree by sketching a model with four discrete types of car. (This sketch is contained in the passage beginning 'Suppose . . . '.) In the four-types model, there is a market in bad used cars but not in good ones. However, this model is not developed in any detail; it serves as a kind of appetizer for the main model, in which no trade takes place at all.
4. As a result of presenting this paper, I have discovered that Schelling's model is much more widely known and admired than I had imagined. It has not had the obvious influence on economics that Akerlof's paper has, but it clearly appeals to methodologically-inclined economists.
5. In passing, I must record my puzzlement at the two-way classification of 'colours' or 'races' which seems to be a social fact in America, despite the continuity of the actual spectra of skin colour, hair type and other supposed racial markers. The convention, I take it, is that anyone of mixed African and European parentage, whatever that mix, is black unless he or she can 'pass' as pure European.
6. When I have presented this paper, I have been surprised at how many economists are inclined towards this interpretation.
7. Arrow (1951: 4–5) hints at this interpretation when, as part of the introduction to his presentation of the theorem, he says that welfare economists need to check that the value judgements they invoke are mutually compatible. He goes on: 'Bergson considers it possible to establish an ordering of social states which is based on the indifference maps of individuals, and Samuelson has agreed'. Arrow's form of social choice theory investigates whether this is indeed possible.
8. This interpretation of Akerlof's model was suggested to me by Daniel Hausman. Hausman also suggested the 'counter-example' interpretation of Schelling's model, discussed in the next paragraph.
9. Here I am using 'story' in the sense which McCloskey (1983: 505) correctly identifies as standard usage among economic theorists: 'an extended example of the economic reasoning underlying the mathematics [of a theory], often a simplified version of the situation in the real world that the mathematics is meant to characterize'. Gibbard and Varian (1978) use 'story' in a similar way (see section 6). Morgan (1997) has a quite different concept of a story. For Morgan, models are inert mechanisms which need to be 'cranked' by some external event

in order to set them in motion; a story is a description of that event and of how its impact is transmitted through the model. Morgan's approach conflates two distinctions – static/dynamic and model/story – which I prefer to keep separate.

10. Early astronomy provides a classic example of the conflict between instrumentalism and realism. The only available observations were of the movements of points and areas of light across the sky. Highly accurate predictions of these movements could be made by using theories based on apparently fantastic and (at the time) completely unverifiable assumptions about how the workings of the universe might look, viewed from outside. With hindsight, we know that some of these fantastic assumptions proved to be true (which supports realism), while others proved false (which supports instrumentalism).
11. The idea that there might be some value in predicting the consumption decisions of individual consumers would perhaps not occur to an economist in the 1950s or 1960s, when the instrumentalist defence of neoclassical theory was most popular. At that time, there were no practicable means to collect or to analyse individual-level data. Developments in retailing and in information technology are now opening up the possibility of making profitable use of predictions about the decisions of individual consumers.
12. Hausman adds the qualification that 'a great deal of theoretical work in economics is concerned with conceptual exploration, not with empirical theorizing' (p. 221). In section 4, I considered and rejected the suggestion that Akerlof's and Schelling's models could be interpreted as conceptual explanation.
13. The parallel between models and experiments is explored in detail by Guala (1999).
14. This interpretation was suggested to me by Maarten Janssen.
15. Cartwright (1998) explores the role of this kind of reasoning in Mill's scientific method.
16. There is an analogy in experimental method. Think of how experimental biologists use fruit flies to test and refine hypotheses about biological evolution. The hypotheses in which the biologists are interested are intended to apply to many species other than fruit flies – sometimes, for example, to humans. Fruit flies are used because they are easy to keep in the laboratory and breed very quickly. But fruit flies are not simplified versions of humans, arrived at by isolating certain key features. Rather, the biologist's claim is that certain fundamental evolutionary mechanisms are common to humans and fruit flies.
17. Akerlof and Schelling are perhaps atypical in that they are satisfied to present simple, imaginative models, leaving it to the technicians of economic theory to produce the generalizations. In contrast, most theorists feel compelled to present their models in the most general form they can. If I am right about the importance of stripping down a model in order to judge how generalizable it is, it is at least arguable that Akerlof's and Schelling's way of presenting models is the more informative.
18. The 'grue' problem discovered by Nelson Goodman (1954) is particularly significant – and intractable.
19. For what it is worth, I am inclined to agree with David Hume's (1740, Book 1, Part 3, pp. 69–179) original diagnosis: that induction is grounded in associations of ideas that the human mind finds natural. If that diagnosis is correct, the

concepts of similarity which underpin inductive reasoning may be capable of being explained in psychological terms, but not of being justified as rational.

20. Notice that one implication of thinking in this way is that regularities within the real world (here, across cities which in many respects are very different from one another) can give us grounds for greater confidence in inductive inferences from a model to the real world. The fact that racial segregation is common to so many different cities suggests that its causes are not to be found in any of those dimensions on which they can be differentiated.
21. I still recall the deep impression made on me as a teenager by Stan Barstow's *A Kind of Loving*. The main character of this classic of northern English realistic fiction is a very ordinary young man who gets his girlfriend pregnant and is then pushed into an unwanted marriage. Reading this book, I gained a vivid sense of the possible consequences for me of actions that I could imagine myself taking.

References

Akerlof, G. A. (1970) 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics* 84: 488–500.

Arrow, K. J. (1963) *Social Choice and Individual Values*, 2nd edn, New Haven, CT: Yale University Press. (1st edn 1951.)

Blaug, M. (1997) 'Ugly currents in modern economics', paper presented at conference *Fact or Fiction? Perspectives on Realism and Economics*, Erasmus University, Rotterdam, November 1997, and in Uskali Mäki (ed.) *Fact and Fiction. Foundational Perspectives on Economics and the Economy*, forthcoming.

Cartwright, N. (1998) 'Capacities', forthcoming in *The Handbook of Methodology*, Aldershot: Edward Elgar.

Gibbard, A. and Varian, H. (1978) 'Economic Models', *Journal of Philosophy* 75: 664–677.

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

Guala, F. (1999) 'Economics and the Laboratory', Ph.D thesis, London School of Economics and Political Science.

Hausman, D. M. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.

Hume, D. (1740) *A Treatise of Human Nature*, page references to 1978 edn, Oxford: Clarendon Press.

McCloskey, D. (1983) 'The Rhetoric of Economies', *Journal of Economic Literature* 21: 481–517.

Mäki, U. (1992) 'On the Method of Isolation in Economics', *Poznań Studies in the Philosophy of the Sciences and the Humanities* 26: 316–351.

Mäki, U. (1994) 'Isolation, Idealization and Truth in Economics', *Poznań Studies in the Philosophy of the Sciences and the Humanities* 38: 147–168.

Mill, J. S. (1843) *A System of Logic*, page references to 1967 edn, London: Longman.

Morgan, M. S. (1997) 'Models, Stories and the Economic World', paper presentepresented at conference *Fact or Fiction? Perspectives on Realism and Economics*, Erasmus University, Rotterdam, November 1997, and in Uskali Mäki (ed.) *Fact and Fiction. Foundational Perspectives on Economics and the Economy*, forthcoming.

Schelling, T. C. (1978) *Micromotives and Macrobehaviour*, New York: Norton.

Selected Bibliography of Books on Economic Methodology*

Alt, J., M. Levi, and E.Ostrom, eds. *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science.* New York: Russell Sage Foundation, 1999.

Arrow, K. *Social Choice and Individual Values*. New York: Wiley, 1963.

Aulin, A. *The Origins of Economic Growth: The Fundamental Interaction between Material and Nonmaterial Values*. Heidelberg and New York: Springer, 1997.

Backhouse, R. *Truth and Progress in Economic Knowledge (Advances in Economic Methodology)*. Cheltenham: Edward Elgar, 1997.

———. *Explorations in Economic Methodology*. London: Routledge, 1998.

Backhouse, R., ed. *New Directions in Economic Methodology*. London: Routledge, 1994.

———. *Methodology of Economics*. London: Routledge, 1997.

Backhouse, R., D. Hausman, and U. Mäki, eds. *Economics and Methodology: Crossing Boundaries*. London: Palgrave Macmillan, 1998.

Backhouse, R., and B. Bateman, eds. *The Cambridge Companion to Keynes*. Cambridge: Cambridge University Press, 2006.

Balak, B. *McCloskey's Rhetoric: Discourse Ethics in Economics*. London: Routledge, 2006.

Balzer, W., and B. Hamminga, eds. *Philosophy of Economics*. Dordrecht: Kluwer-Nijhoff, 1989.

Barker, D., and E. Kuiper, eds. *Toward a Feminist Philosophy of Economics*. London: Routledge, 2003.

Barnes, T. *Logics of Dislocation: Models, Metaphors, and Meanings of Economic Space*. New York and London: Guilford Press, 1996.

Becker, G. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, 1976.

* To provide a comprehensive bibliography of books and articles is impossible. There is a useful bibliography of works on economic methodology through 1988 (Redman 1989). Essays on economic methodology are indexed in *The Journal of Economic Literature* and in the *Index of Economic Articles in Journal and Collective Volumes* under the number 036 before 1991 and under the number B4 since then. *Economics and Philosophy* and *The Journal of Economic Methodology* are the best places to start for recent essays on economic methodology. Books on economics and ethics and on the theory of rationality are included in this Bibliography only if they have some substantial methodological content.

Bell, D., and I. Kristol, eds. *The Crisis in Economic Theory*. New York: Basic Books, 1981.

Bhaskar, R., M. Archer, A. Collier, T. Lawson, and A. Norrie, eds. *Critical Realism*. London: Routledge, 1998.

Birner, J. *Strategies and Programmes in Capital Theory: A Contribution to the Methodology of Theory Development*. Dissertation, University of Amsterdam, 1990.

Blaug, M. *The Cambridge Revolution. Success or Failure?* London: Institute of Economic Affairs, 1975.

———. *Economic Theory in Retrospect*. 3rd ed. Cambridge: Cambridge University Press, 1978.

———. *A Methodological Appraisal of Marxian Economics*. Amsterdam: North-Holland, 1980a.

———. *The Methodology of Economics: Or How Economists Explain*. Cambridge: Cambridge University Press, 1980b; 2nd ed. 1992.

Boehm, S. *et al.*, eds. *Is There Progress in Economics? Knowledge, Truth and the History of Economic Thought*. Cheltenham: Edward Elgar, 2002.

Boland, L. *The Foundations of Economic Method*. London: Allen & Unwin, 1982.

———. *Methodology for a New Microeconomics*. Boston: Allen & Unwin, 1986.

———. *The Methodology of Economic Model Building: Methodology after Samuelson*. London: Routledge, 1989.

———. *Critical Economic Methodology: A Personal Odyssey*. London: Routledge, 1997.

Bonar, J. *Philosophy and Political Economy* (1893). rpt. London: Allen & Unwin, 1967.

Boulding, K. *Economics as a Science*. New York: McGraw-Hill, 1970.

Boylan, T., and P. O'Gorman. *Beyond Rhetoric and Realism in Economics: Towards a Reformulation of Economic Methodology*. London: Routledge, 1995.

Brennan, G., H. Kliemt, and R. Tollison, eds. *Method and Morals in Constitutional Economics: Essays in Honor of James M. Buchanan*. New York: Springer, 2002.

Bromley, D. *Sufficient Reason: Volitional Pragmatism and the Meaning of Economic Institutions*. Princeton: Princeton University Press, 2006.

Cairnes, J. *The Character and Logical Method of Political Economy*. 2nd ed. (1875). rpt. New York: A. M. Kelley, 1965.

Caldwell, B. *Beyond Positivism: Economic Methodology in the Twentieth Century*. London: Allen & Unwin, 1982.

Caldwell, B., ed. *Appraisal and Criticism in Economics*. London: Allen & Unwin, 1984.

———. *The Philosophy and Methodology of Economics*. Cheltenham: Edward Elgar, 1993.

Caplin, A., and A. Schotter, eds. *Perspectives on the Future of Economics: Positive and Normative Foundations*. Oxford: Oxford University Press, 2007.

Cory, G. *The Consilient Brain: The Bioneurological Basis of Economics, Society, and Politics*. Dordrecht: Kluwer, 2004.

Cullenberg, S., J. Amariglio, and F. Ruccio, eds. *Postmodernism, Economics, and Knowledge*. London: Routledge, 2001.

Dahiya, S., ed. *The Current State of Economic Science*. Vol. 1. *Teaching of Economics, Methodology, Mathematical and Quantitative Methods*. Rohtak, India: Spellbound, 1999.

D'Autume, A., and J. Cartelier, eds. *Is Economics Becoming a Hard Science?* Cheltenham: Edward Elgar, 1997.

Davis, J. *Theory of the Individual in Economics*. London: Routledge, 2003.

Davis, J., ed. *Recent Developments in Economic Methodology*. Cheltenham: Edward Elgar, 2006.

Davis, J., D. W. Hands, and U. Mäki, eds. *The Handbook of Economic Methodology*. Cheltenham: Edward Elgar, 1998.

Davis, J., A. Marciano, and J. Runde, eds. *The Elgar Companion to Economic and Philosophy*. Cheltenham: Edward Elgar, 2004.

Delorme, R., and K. Dopfer, eds. *The Political Economy of Diversity: Evolutionary Perspectives on Economic Order and Disorder*. Aldershot: Ashgate, 1994.

de Marchi, N., ed. *The Popperian Legacy in Economics*. Cambridge: Cambridge University Press, 1988.

______. *Post-Popperian Methodology of Economics: Recovering Practice*. Boston: Kluwer, 1992.

De Marchi, D., and M. Blaug, eds. *Appraising Modern Economics: Studies in the Methodology of Scientific Research Programs*. Cheltenham: Edward Elgar, 1991.

Dennis, K., ed. *Rationality in Economics: Alternative Perspectives*. Dordrecht: Kluwer, 1998.

Diesing, P. *Science and Ideology in the Policy Sciences*. New York: Aldine, 1982.

Dolan, E., ed. *The Foundations of Modern Austrian Economics*. Kansas City, MO: Sheed & Ward, 1976.

Dopfer, K., ed. *Economics and Philosophy: The Evolutionary Foundations of Economics*. Cambridge: Cambridge University Press, 2005.

Dow, S. *Macroeconomic Thought: A Methodology Approach*. Oxford: Blackwell, 1985.

______. *Economic Methodology: An Inquiry*. Oxford: Oxford University Press, 2002.

Dyke, C. *Philosophy of Economics*. Englewood Cliffs, NJ: Prentice Hall, 1981.

Eichner, A., ed. *Why Economics Is Not Yet a Science*. Armonk, NY: M. E. Sharpe, pp. 205–41.

Elster, J. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press, 1979.

______. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press, 1983.

______. *Making Sense of Marx*. Cambridge: Cambridge University Press, 1985.

Elster, J., and J. Roemer, eds. *Interpersonal Comparisons of Well-Being*. Cambridge: Cambridge University Press, 1991.

Etzioni, A. *The Moral Dimension: Toward a New Economics*. New York: Macmillan, 1988.

Fisher, R. *The Logic of Economic Discovery: Neoclassical Economics and the Marginal Revolution*. New York: New York University Press, 1986.

Fleetwood, S., ed. *Critical Realism in Economics: Development and Debate*. London: Routledge, 1999.

Foldvary, F., ed. *Beyond Neoclassical Economics: Heterodox Approaches to Economic Theory*. Cheltenham: Edward Elgar, 1996.

Fox, G. *Reason and Reality in the Methodologies of Economics: An Introduction*. Cheltenham: Edward Elgar, 1997.

Frankfurter, G., and E. McGoun, eds. *From Individualism to the Individual: Ideology and Inquiry in Financial Economics*. Aldershot: Ashgate, 2002.

Fraser, L. *Economic Thought and Language. A Critique of Some Fundamental Concepts*. London: A & C Black, 1937.

Frey, B. *Economics as a Science of Human Behaviour: Towards a New Social Science Paradigm*. 2nd ed. Dordrecht: Kluwer, 1999.

Friedman, M. *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953.

Fullbrook, E. *A Guide to What's Wrong with Economics*. New York: Anthem Press, 2004.

Fullbrook, E., ed. *Intersubjectivity in Economics: Agents and Structures*. London: Routledge, 2002.

Gani, M. *Foundations of Economic Science*. Dhaka, Bangladesh and Ontario: Scholars, 2003.

Grönkvist, U. *Economic Methodology: Patterns of Reasoning and the Structure of Theories*. Lund: University of Lund, 1992.

Grossbard-Shechtman, S., and C. Clague, eds. *The Expansion of Economics: Toward a More Inclusive Social Science*. Armonk, NY: M. E. Sharpe, 2002.

Guala, F. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press, 2005.

Gustafsson, B., C. Knudsen, and U. Mäki, eds. *Rationality, Institutions and Economic Methodology*. London: Routledge, 1993.

Hahn, F., and M. Hollis. *Philosophy and Economic Theory*. Oxford: Oxford University Press, 1979.

Hamminga, B. *Neoclassical Theory Structure and Theory Development: An Empirical-Philosophical Case Study Concerning the Theory of International Trade*. Boston: Springer, 1983.

Hamminga, B., and N. DeMarchi, eds. *Idealization in Economics*. Amsterdam: Rodopi, 1994.

Hanappi, H. *Evolutionary Economics: The Evolutionary Revolution in the Social Sciences*. Aldershot: Ashgate, 1994.

Hands, D. W. *Testing, Rationality and Progress*. Totowa, NJ: Rowman and Littlefield, 1992.

______. *Reflection without Rules: Economic Methodology and Contemporary Science Theory*. Cambridge: Cambridge University Press, 2001.

Hausman, D. *Capital, Profits and Prices: An Essay in the Philosophy of Economics*. New York: Columbia University Press, 1981.

______. *Essays on Philosophy and Economic Methodology*. Cambridge: Cambridge University Press, 1992a.

______. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press, 1992b.

Hausman, D., and McPherson, M. *Economic Analysis and Moral Philosophy*. Cambridge: Cambridge University Press, 1996.

______. *Economic Analysis, Moral Philosophy, and Public Policy*. Cambridge: Cambridge University Press, 2006.

Hayek, F. *The Counter-Revolution of Science: Studies in the Abuse of Reason*. Glencoe, IL: Free Press, 1952.

Henderson, W., T. Dudley-Evans, and R. Backhouse, eds. *Economics and Language*. London: Routledge, 1993.

Hicks, J. *Causality in Economics*. New York: Basic Books, 1979.

Hirsch, A., and N. de Marchi. *Milton Friedman: Economics in Theory and Practice*. Ann Arbor: University of Michigan Press, 1990.

Hirsch, F. *The Social Limits to Growth*. Cambridge, MA: Harvard University Press, 1976.
Hodgson, B. *Economics as Moral Science*. Heidelberg and New York: Springer, 2001.
Hodgson, G. *The Evolution of Institutional Economics*. London: Routledge, 2004.
Hollis, M., and E. Nell. *Rational Economic Man: A Philosophical Critique of Neo-Classical Economics*. London: Cambridge University Press, 1975.
Hood, W., and T. Koopmans, eds. *Studies in Econometric Method*. New York: John Wiley, 1953.
Hook, S., ed. *Human Values and Economic Policy*. New York: New York University Press, 1967.
Hoover, K. *The New Classical Macroeconomics: A Sceptical Inquiry*. Oxford: Basil Blackwell, 1988.
______. *Causality in Macroeconomics*. Cambridge: Cambridge University Press, 2001.
______. *The Methodology of Empirical Macroeconomics*. Cambridge: Cambridge University Press, 2001.
Hoover, K., and S. Sheffrin, eds. *Monetarism and the Methodology of Economics: Essays in Honour of Thomas Mayer*. Aldershot: Ashgate, 1995.
Humphries, J., ed. *Gender and Economics*. Aldershot: Edward Elgar, 1995.
Hutchison, T. *The Significance and Basic Postulates of Economics* (1938). rpt. with a new Preface. New York: A. M. Kelley, 1960.
______. *Knowledge and Ignorance in Economics*. Chicago: University of Chicago Press, 1977.
______. *On Revolutions and Progress in Economic Knowledge*. Cambridge: Cambridge University Press, 1978.
______. *The Politics and Philosophy of Economics: Marxians, Keynesians and Austrians*. Oxford: Blackwell, 1981.
______. *The Uses and Abuses of Economics: Contentious Essays on History and Method*. London: Routledge, 1994.
______. *On the Methodology of Economics and the Formalist Revolution*. Cheltenham: Edward Elgar, 2000.
Kamarck, A. *Economics and the Real World*. Philadelphia: University of Pennsylvania Press, 1983.
______. *Economics for the Twenty-First Century: The Economics of the Economist-Fox*. Aldershot: Ashgate, 2001.
Katzner, D. *Time, Ignorance, and Uncertainty in Economic Models*. Ann Arbor: University of Michigan Press, 1998.
Katouzian, H. *Ideology and Method in Economics*. New York: New York University Press, 1980.
Kaufmann, F. *Methodology of the Social Sciences*. London: Oxford University Press, 1944.
Keen, S. *Debunking Economics: The Naked Emperor of the Social Sciences*. New York: St. Martin's Press, 2001.
Keynes, J. N. *The Scope and Method of Political Economy* (4th ed. 1917). rpt. New York: A. M. Kelley, 1955.
Kincaid, H. *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press, 1996.
Kirzner, I. *The Economic Point of View*. 2nd ed. Kansas City, MO: Sheed & Ward, 1976.
Kirzner, I., ed. *Method, Process and Austrian Economics: Essays in Honour of Ludwig von Mises*. Lexington, MA: D. C. Heath, 1982.

Klamer, A. *Conversations with Economists: New Classical Economists and Opponents Speak Out on the Current Controversy in Macroeconomics.* Totowa, NJ: Rowman and Allanheld, 1984.

Klamer, A., D. McCloskey, and R. Solow, eds. *The Consequences of Economic Rhetoric.* New York: Cambridge University Press, 1988.

Klant, J. *The Rules of the Game.* Cambridge: Cambridge University Press, 1984.

Koopmans, T. *Three Essays on the State of Economic Science.* New York: McGraw-Hill, 1956.

Kornai, J. *Anti-Equilibrium: On Economic Systems Theory and the Tasks of Research.* Amsterdam: North Holland, 1971.

Koslowski, P., ed. *Economics and Philosophy*. Tübingen: J.C.B. Mohr, 1985.

______. *Sociobiology and Bioeconomics: The Theory of Evolution in Biological and Economic Theory*. Heidelberg and New York: Springer, 1999.

Krupp, S., ed. *The Structure of Economic Science.* Englewood Cliffs: Prentice Hall, 1966.

Lapavitsas, C. *Social Foundations of Markets, Money and Credit.* London: Routledge, 2003.

Latsis, S., ed. *Method and Appraisal in Economics.* Cambridge: Cambridge University Press, 1976.

Lavoie, D., ed. *Economics and Hermeneutics.* London: Routledge, 1991.

Lawson, T. *Economics and Reality*. London: Routledge, 1997.

______. *Reorienting Economics.* London: Routledge, 2003.

Lawson, T., and H. Pesaran. *Keynes' Economics: Methodological Issues.* Beckenham, UK: Croom Helm, 1985.

Leibenstein, H. *Beyond Economic Man: A New Foundation for Economics.* Cambridge, MA: Harvard University Press, 1976.

Levine, A., E. Sober, and E. Wright. *Reconstructing Marxism.* London: Verso, 1992.

Lewis, P., ed. *Transforming Economics.* London: Routledge, 2004.

Little, D., ed. *On the Reliability of Economic Models: Essays in the Philosophy of Economics.* Boston: Kluwer, 1993.

Loasby, B. *Choice, Complexity and Ignorance.* Cambridge: Cambridge University Press, 1976.

______. *The Mind and Method of the Economist: A Critical Appraisal of Major Economists in the 20th Century*. Cheltenham: Edward Elgar, 1989.

Lowe, A. *On Economic Knowledge. Toward a Science of Political Economics.* New York: Harper & Row, 1965.

McClelland, P. *Causal Explanation and Model Building in History, Economics and the New Economic History*. Ithaca, NY: Cornell University Press, 1975.

McClennen, E. *Rationality and Dynamic Choice: Foundational Explorations.* Cambridge: Cambridge University Press, 1990.

McCloskey, D. *The Rhetoric of Economics.* Madison: University of Wisconsin Press, 1985.

______. *If You're So Smart: The Narrative of Economic Expertise.* Chicago: University of Chicago Press, 1990.

______. *Knowledge and Persuasion in Economics.* Cambridge: Cambridge University Press, 1994.

______. *The Vices of Economists – The Virtues of the Bourgeoisie.* Amsterdam: Amsterdam University Press, 1996.

Machlup, F. *Essays on Economic Semantics*. Ed. M. Miller. Englewood Cliffs: Prentice Hall, 1963.

———. *Methodology of Economics and Other Social Sciences*. New York: Academic Press, 1978.

MacKay, A. *Arrow's Theorem: The Paradox of Social Choice. A Case Study in the Philosophy of Economics*. New Haven, CT: Yale University Press, 1980.

McKenzie, R. *The Limits of Economic Science*. Boston: Kluwer, 1983.

Marr, W., and B. Raj, eds. *How Economists Explain: A Reader in Methodology*. Lanham, MD: University Press of America, 1983.

Mayer, T. *Truth versus Precision in Economics*. Cheltenham: Edward Elgar, 1993.

Mayer, T., K. Hoover, and S. Sheffrin, eds. *Monetarism and the Methodology of Economics: Essays in Honour of Thomas Mayer*. Cheltenham: Edward Elgar, 1995.

Mäki, U. *Studies in Realism and Explanation in Economics*. Helsinki: Suomalainen Tiedeakatemia, 1990.

———. *Realism and Economic Methodology*. London: Routledge, 2007.

Mäki, U., ed. *The Economic World View: Studies in the Ontology of Economics*. Cambridge: Cambridge University Press, 2001.

———. *Fact and Fiction in Economics: Models, Realism and Social Construction*. Cambridge: Cambridge University Press, 2003.

Medema, S., ed. *Coasean Economics: Law and Economics and the New Institutional Economics*. Dordrecht: Kluwer, 1998.

Meidinger, C. *Science Économique: Questions de Méthode*. Paris: Vuibert, 1994.

Menger, C. *Problems of Economics and Sociology* (1883). Ed. L. Schneider, tr. F. Nock. Urbana: University of Illinois Press, 1963.

Mill, J. S. *A System of Logic* (1843). rpt. London: Longmans, Green & Co., 1949.

Mirowski, P. *Against Mechanism: Protecting Economics from Science*. Totowa, NJ: Rowman and Littlefield, 1988.

———. *More Heat than Light*. Cambridge: Cambridge University Press, 1990.

———. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press, 2001.

———. *The Effortless Economy of Science?* Durham, NC: Duke University Press, 2004.

Mirowski, P., ed. *The Reconstruction of Economic Theory*. Boston: Kluwer, 1986.

Mises, L. von. *Human Action. A Treatise on Economics*. New Haven, CT: Yale University Press, 1949.

———. *Epistemological Problems of Economics*. tr. G. Reisman. New York: New York University Press, 1981.

———. *The Ultimate Foundation of Economic Science: An Essay on Method*. 2nd ed. Kansas City, MO: Sheed Andrews, 1978.

Morgan, M. *History of Econometric Ideas*. Cambridge: Cambridge University Press, 1990.

Mosini, V. ed. *Equilibrium in Economics*. London: Routledge, 2007.

Mulberg, J. *Social Limits to Economic Theory*. London: Routledge, 1995.

Myrdal, G. *The Political Element in the Development of Economic Thought*. Tr. P. Streeten. Cambridge, MA: Harvard University Press, 1955.

———. *Objectivity in Social Research*. London: Duckworth, 1970.

Nelson, R. *Economics as Religion: From Samuelson to Chicago and Beyond*. College Station: Pennsylvania State University Press, 2001.

Nelson, R., and S. Winter. *An Evolutionary Theory of Economic Change.* Cambridge, MA: Harvard University Press, 1982.
Neuberg, L. *Conceptual Anomalies in Economics.* Cambridge: Cambridge University Press, 1988.
O'Sullivan, P. *Economic Methodology and Freedom to Choose.* London: Allen & Unwin, 1987.
Oakley, A. *Reconstructing Economic Theory: The Problem of Human Agency*. Cheltenham: Edward Elgar, 2002.
Ochangco, A. *Rationality in Economic Thought: Methodological Ideas on the History of Political Economy*. Cheltenham: Edward Elgar, 1999.
Omerod, P. *The Death of Economics.* New York: Wiley, 1997.
Pani, N. *Inclusive Economics: Gandhian Method and Contemporary Policy*. New Delhi and London: Sage Publications, 2001.
Papandreou, A. *Economics as a Science.* Chicago: Lippincott, 1958.
Parson, T. *The Structure of Social Action.* New York: McGraw-Hill, 1937, vol. 2.
Pheby, J. *Methodology and Economics: A Critical Introduction.* London: Macmillan, 1988.
Pitt, J., ed. *Philosophy in Economics.* Dordrecht: Reidel, 1981.
Rappaport, S. *Models and Reality in Economics.* Cheltenham: Edward Elgar, 1998.
Reder, M. *Economics: The Culture of a Controversial Science.* Chicago: University of Chicago Press, 1999.
Redman, D. *Economic Methodology: A Bibliography with References to Works in the Philosophy of Science, 1860–1988.* New York: Greenwood Press, 1989.
______. *Economics and the Philosophy of Science.* Oxford: Oxford University Press, 1990.
______. *The Rise of Political Economy as a Science: Methodology and the Classical Economists.* Cambridge, MA: MIT Press, 1997.
Reiss, J. *Error in Economics: Towards a More Evidence-Based Methodology*. London: Routledge, 2007.
Robbins, L. *An Essay on the Nature and Significance of Economic Science.* 2nd ed. London: Macmillan, 1935. 3rd ed. 1983.
Robinson, J. *Economic Philosophy*. Chicago: Aldine, 1962.
Rosenberg, A. *Microeconomic Laws: A Philosophical Analysis.* Pittsburgh: University of Pittsburgh Press, 1976.
______. *Economics – Mathematical Politics or Science of Diminishing Returns.* Chicago: University of Chicago Press, 1992.
Ross, D. *Economic Theory and Cognitive Science: Microexplanation.* Cambridge, MA: MIT Press, 2005.
Rothbard, M. *The Logic of Action I: Method, Money, and the Austrian School.* Cheltenham: Edward Elgar, 1997.
Rothschild, K. *Ethics and Economic Theory*. Cheltenham: Edward Elgar, 1993.
______. *Economic Method, Theory and Policy: Selected Essays of Kurt W. Rothschild.* Ed. J. King. Aldershot: Ashgate, 1995.
Roy, S. *Philosophy of Economics: On the Scope of Reason in Economic Inquiry*. London: Routledge, 1991.
Ruccio, D., and J. Amariglio. *Postmodern Moments in Modern Economics.* Princeton: Princeton University Press, 2003.
Runde, J., and S. Mizuhara, eds. *The Philosophy of Keynes's Economics: Probability, Uncertainty and Convention.* London: Routledge, 2003.

Salanti, A., and E. Screpanti, eds. *Pluralism in Economics: New Perspectives in History and Methodology*. Cheltenham: Edward Elgar, 1997.

Samuels, W., ed. *The Methodology of Economic Thought: Critical Papers from the Journal of Economic Thought [Issues]*. New Brunswick, NJ: Transaction Books, 1980.

———. *History and Methodology of Economics*. Greenwich, CT: JAI Press, 1987.

———. *Economics as Discourse*. Boston: Kluwer, 1990.

Samuels, W., and J. Biddle, eds. *Research in the History of Economic Thought and Methodology*. Vol. 16. London: JAI Press, 1998.

Samuelson, P. *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press, 1947.

Sassower, R. *Philosophy of Economics: A Critique of Demarcation*. Lanham, MD: University Press of America, 1985.

Schoeffler, S. *The Failures of Economics: A Diagnostic Study*. Cambridge, MA: Harvard University Press, 1955.

Schrader, D. *The Corporation as Anomaly*. Cambridge: Cambridge University Press, 1992.

Schumpeter, J. *History of Economic Analysis*. New York: Oxford University Press, 1954.

Sen, A. *On Ethics and Economics*. Oxford: Blackwell, 1987.

Senior, N. *Outline of the Science of Political Economy* (1836). rpt. New York: A. M. Kelley, 1965.

Shackle, G. *Epistemics and Economics: A Critique of Economic Doctrines*. Cambridge: Cambridge University Press, 1972.

Shrader-Frechette, K. *Science Policy, Ethics, and Economic Methodology: Some Problems of Technology Assessment and Environmental-Impact Analysis*. Dordrecht: D. Reidel, 1984.

Sidgwick, H. *The Scope and Method of Economic Science* (1885). rpt. New York: A. M. Kelley, 1968.

Simon, H. *Models of Man: Social and Rational*. New York: Wiley, 1957.

———. *An Empirically Based Microeconomics*. Cambridge: Cambridge University Press, 1997.

Smyth, R., ed. *Essays in Economic Method*. London: Duckworth, 1962.

Sowell, T. *Knowledge and Decisions*. New York: Basic Books, 1980.

Staveren, I. van. *Values of Economics*. London: Taylor and Francis, 2005.

Stegmüller, W., W. Balzer, and W. Spohn, eds. *Philosophy of Economics: Proceedings, Munich, July 1981*. New York: Springer, 1982.

Stewart, I. *Reasoning and Method in Economics. An Introduction to Economic Methodology*. London: McGraw-Hill, 1979.

Stigum, B. *Econometrics and the Philosophy of Economics: Theory – Data Confrontations in Economics*. Princeton: Princeton University Press, 2003.

Swedberg, R. *Economics and Sociology – Redefining Their Boundaries: Conversations with Economists and Sociologists*. Princeton: Princeton University Press, 1990.

Verdon, M. *Keynes and the "Classics": A Study in Language, Epistemology and Mistaken Identities*. London: Routledge, 1996.

Vickers, D. *The Tyranny of the Market: A Critique of Theoretical Foundations*. Ann Arbor: University of Michigan Press, 1995.

———. *Economics and Ethics: An Introduction to Theory, Institutions, and Policy*. London: Greenwood, Praeger, 1997.

Vromen, J. *Economic Evolution*. London: Routledge, 1995.

Ward, B. *What's Wrong with Economics?* New York: Basic Books, 1972.

Weber, M. *The Methodology of the Social Sciences*. Tr. and ed. E. Shils and H. Finch. New York: Macmillan, 1949.

Weintraub, E. *General Equilibrium Analysis: Studies in Appraisal*. Cambridge: Cambridge University Press, 1985.

———. *Stabilizing Dynamics*. Cambridge: Cambridge University Press, 1991.

———. *How Economics Became a Mathematical Science*. Durham, NC: Duke University Press, 2002.

Wible, J. *The Economics of Science: Methodology and Epistemology as if Economics Really Mattered*. London: Routledge, 1998.

Wiles, P., and G. and Routh, eds. *What is Political Economy? Eight Perspectives*. Oxford: Basil Blackwell, 1984.

Winston, G., and R. Teichgraeber, eds. *The Boundaries of Economics*. Cambridge: Cambridge University Press, 1988.

Wiseman, J., ed. *Beyond Positive Economics?* London: British Association for the Advancement of Science, 1983.

Wong, S. *The Foundations of Paul Samuelson's Revealed Preference Theory*. London: Routledge, 1978.

Yuengert, A. *The Boundaries of Technique: Ordering Positive and Normative Concerns in Economic Research*. Lanham, MD: Lexington Books, 2004.

Zein-Elabdin, E., and S. Charusheela, eds. *Postcolonialism Meets Economics*. London: Routledge, 2003.