Milton Friedman
"The Methodology of Positive Economics"
In *Essays In Positive Economics*
(Chicago: Univ. of Chicago Press, 1966), pp. 3-16, 30-43.

*The Methodology of Positive Economics**

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 paper 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 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.

1. 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

* 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. 455-57.

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

1. (London: Macmillan 4 Co., 1891), pp. 34-35 and 46.

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

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.

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

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 Godel'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.

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 "economies 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

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.

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

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.* 56-57, 90-91.

6. See 'Lange on Price Flexibility and Employment: A Methodological Criticism," *infra, pp.* 282-89.

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

7. "The Marshallian Demand Curve," *infra*, p. 57.

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

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. 282-83.

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 inflation 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

9. See "Lange on Price Flexibility and Employment," infra, passim.

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

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. 465-67.

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 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

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.), *Statistical 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.*

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

12. The converse of the proposition does not- of course hold: assumptions that are unrealistic (in this sense) do not guarantee a significant theory.

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

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,

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.

"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 also 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.

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

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.

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 of an alternative hypothesis, on the other.

22. See H. D. Henderson, "The Significance of the Rate of Interest," *Oxford Economic Papers*, No. I (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. I (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 Re*view, XXXIX (May, 1949), 198-99.

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 multi-process 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

23. Oliver, op. cit, p. 382.

24. Lester, "Shortcomings of Marginal Analysis for Wage-Employment Problems," op. cit., P. 75.

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

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 hypothesis as 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.

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, unsuccesful 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

27. Principles of Political Economy (Ashley ed.; Longmans, Green & Co., 1929), p. 436.

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 ni~ 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

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.

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 $S = 1/2gt^2$ 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

31. *Principles*, p. 100.

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 Marlborough.[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,

32. Quoted parts from ibid.

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

33. E. H. Chamberlin, *The Theory of Monopolistic Competition (6tb* e.d.; Cambridge: Harvard University. Press, *1950);* Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan & Co., *1933).*

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." 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.

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. 12-24.

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 fundamental 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 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; studied in autobiographies and biographies, not treatises on scientific method; and promoted by maxim and example, not syllogism or theorem.