

Essays in
Positive Economics

BY
MILTON FRIEDMAN

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO AND LONDON

*Acknowledgment is hereby made to the following publishers
for permission to reprint material which originally
appeared over their imprints:*

AMERICAN ECONOMIC REVIEW
RICHARD D. IRWIN, INC.
HARVARD UNIVERSITY PRESS

ISBN: 0-226-26403-3 (paperbound)
LCN: 53-3533

THE UNIVERSITY OF CHICAGO PRESS, CHICAGO, 60637
The University of Chicago Press, Ltd., London

Copyright 1953 by The University of Chicago. All rights reserved
Published 1953. Printed in the United States of America

01 00 99 98 97 96 13 14 15 16 17

© The paper used in this publication meets the minimum requirements
of the American National Standard for Information Sciences—Permanence
of Paper for Printed Library Materials, ANSI Z39.48—1984.

tinguishing 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 pre-

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.

ceding 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 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 tautol-

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

ogies 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.

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 $s = \frac{1}{2}gt^2$, 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

"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.

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.

This 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 $s = \frac{1}{2} gt^2$ 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 $s = \frac{1}{2} gt^2$. 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 a 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

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.

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,

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 distributions 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.

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 memo-

randums, 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.