

## On the Career of a Microeconomist \*

For as long as I can remember — certainly by my early teens — my desire to be an economist was never in doubt. In retrospect I can see that this was no accident. Both my parents were self-educated immigrants, but educated they were, as few undergraduates are today. Literature, older or current, language, politics and economics were constantly discussed in our house in a most fascinating manner, and I was expected from childhood to participate fully in the discussions. My father with his lower class background (his parents had run a tavern in a small town in Poland) was driven by passionate concern for humanity, and emotion ruled his talk. My mother, by contrast, coming from a line of Jewish Lithuanian intellectuals, epitomized logic and careful reasoning in pursuit of the same objectives. Both parents, particularly my father, were avid Marxists. The combination was irresistible. I was infected by their interests and their concerns. My reading provided a sampling of Marx's logical convolutions which, combined with tales of the adventure of the buccaneers of 19th century business — Morgan, Vanderbilt, Gould, Rockefeller and others — sealed my fascination with the subject. Well before I entered college I had begun to read economic history, the works of the classical economists and the writings of Thorstein Veblen.

### Undergraduate Education (1939-1942)

I entered C.C.N.Y. (the College of the City of New York) in 1939 as the great depression was drawing to an end and the threat of war hung over us. C.C.N.Y., then, was an extraordinary institution. There was no tuition charge and the students commuted from home every day

---

\* Contribution to a series of recollections and reflections on professional experiences of distinguished economists. This series opened with the September 1979 issue of this *Review*.

by subway. The ambitious children of impecunious immigrants flocked to it as their ticket from poverty or grinding labor. Never did so many future Nobel prize winners congregate in one place. Passionate activity, discussion and debate flourished everywhere. The dining room was surrounded by alcoves each had been claimed by a permanent discussion group — there was the Trotskyist alcove, the Socialist alcove and one of every other variety. (There was even an alcove at which future comedians or rather, comic social commentators, some of them to become famous later, practiced their craft without let up.)

A number of economists of note emerged from there, Kenneth Arrow, Julius Margolis and Jules Joskow among them. But while C.C.N.Y. at that time had many gifted teachers, the department of economics had very few. It was clear to us students that many of them were thoroughly behind the times, and could not teach us about the work of Keynes, Chamberlin and Joan Robinson which were then at the frontier. Besides, many of them just could not communicate very well. There seemed to us to be little choice, and so we organized our own classes, each specializing in a different field, devouring as much of the relevant literature as we could and then lecturing on it to the others. I was assigned the microeconomics, and suspect that I learned more economics there than ever before or since. This experience has always engendered uncertainty in my views on teaching. Can it be that, at least for some students, what is considered to be "bad" teaching is really teaching in its most effective form, because it forces students to think and learn for themselves? I have long advocated some controlled experiments in which parallel classes are held, some taught in the conventional way by teachers of good reputation while in the others students are asked to fend for themselves, guided only by past examinations, perhaps by a reading list, and with no access to the faculty. It is my conjecture that the first group will perform better on the examination at the end of the semester, but that in an examination five or ten years later the self-taught group will far outdistance the other.

### The Department of Agriculture and Military Service (1942-46)

Graduating in 1942, I spent a few months in Washington working (though it may seem odd) in the Department of Agriculture. The shadow of the receding depression led graduating students to accept with delight virtually any reasonable job offer, and I was considered

very daring by the others in my class when I held out for — and received — a salary of \$ 2,000 a year. Happily the job was not undesirable. It was a sort of think tank for the Department, and many of the group, some of whom I still see, were exceedingly bright and creative persons. The head of the Department was Frederick Waugh, a bright and fatherly figure who then and later made a number of significant contributions to the economic literature. It was under his guidance and that of my young colleagues that I first learned how microtheory can be applied to concrete issues.

Just before leaving for Washington I had been married. We were both very young, but the U.S. had just entered the war and many others were doing the same. I have had many occasions to congratulate myself on the wisdom (or luck) that guided that decision. We have ever since worked (and enjoyed life) together in many areas, most directly in our work on the economics of the performing arts. But I am getting ahead of the story.

My years in the U.S. Army are pertinent only for the correspondence course in linear algebra I took, and, while stationed in Rouen, the mathematics books I was able to buy and devour. There, incidentally I made lifelong friends with a French family whose fabulous cellar, including samples of very great vintage of the century, introduced me to the delights of wine which I still collect. Just after the war in Europe had ended some German prisoners gave me my first lessons in wood sculpture which I now teach at Princeton University. (Before entering the army I had studied drawing and painting at the Art Student's League in New York City and my second main field of study at C.C.N.Y. had been visual art — painting and lithography, particularly.)

After returning from the war there was another brief stint at the Department of Agriculture. But this time the task was very different: the allocation of the U.S. grain resources among the countries of a hungry world. Two prime lessons emerged from this experience — the high costs of the negotiation process and the complexities of the calculations of fairness. The first lesson flowed from the fact that all of the senior and more experienced members of the division were almost fully employed in diplomatic negotiations involving international agencies and other governments. As a result, the actual decisions were, to our astonishment, left to another young man of equal inexperience and myself!

It was not a task to make for an easy conscience. The refrain we heard from virtually every country was the same; we are hungry. If we had one shipload to go either to country A or to country B, what were

we to do if the evidence showed that we had already sent enough to A this month to provide each of its inhabitants more calories than before the war, but that B, whose per capita grain receipts were higher than A's, was nevertheless well below its prewar consumption level? We did not, of course, enjoy the luxury of indecision that is available to pure research.

### The London School of Economics (1946-49)

I had assumed that on my return to civilian life I would begin my postgraduate studies. My application to the London School of Economics was rejected, and I wrote again asking when it would be possible to apply for the following year. As later came out, compassion was still part of the admission process, and it was decided to accept me as a student for the masters degree only. As explained to me later, at LSE they had never heard of C.C.N.Y. and, in any event, my undergraduate record was hardly outstanding. But C.C.N.Y. soon made up for its handicap. The training in fierce debating — quarter neither asked nor given — soon got me (in my view, undeserved) attention in LSE's justly famous seminars. To my amazement, within weeks I had not only been transferred to the Ph.D. program, but was also offered a part-time teaching post which became full time the following year. Lord Robbins afterwards told me that he went back to the records to see where they had made their mistake in rejecting me, and decided that, on the record, they had been right, after all.

LSE was an extraordinarily stimulating place to be. Besides Lionel Robbins I got to know Friedrich Hayek, Arthur Lewis, Nicholas Kaldor and, later, James Meade, among the economists, and Karl Popper and Harold Laski outside the department. Among the students there were Frank Hahn, undergraduate Ralph Turvey and an Australian, David Finch (now at the World Bank) who (in 1948!) was writing a thesis on the theory of stagflation. The weekly Robbins seminar and the stimulation in the common rooms where one met for avid and fruitful conversation were experiences I have never duplicated. The erudition and broad knowledge of Robbins and Laski revealed what humanistic learning can be. Popper, who had not yet reacquired any professional logician colleagues, was prepared to report his latest derivations

frequently to a new assistant lecturer who had taken a course or two in formal logic at C.C.N.Y. Hahn, Turvey and other young people were sources of a continuing flow of new and fruitful ideas. Meade and Hayek were also full of ideas which they constantly tried out on the delighted newcomers.

Because few English faculty members had anything like a Ph.D. and did not take the degree seriously, it was possible for me to write my dissertation at the same time that I held a full time faculty position. I had already planned my dissertation during my military service, and Lionel Robbins agreed to be its supervisor. He never begrudged me time or advice, which was always very helpful. The thesis, later published as *Welfare Economics and the Theory of the State*, took off from the Marshall-Pigou theory of externalities, then a neglected subject widely considered to be of minor importance. It sought to generalize the idea to as diverse a set of subjects as the behavior of competitors, the difficulty of cartel formation and the theory of inflation (all of which we were already characterizing as a prisoners' dilemma problem in the new fangled theory of games).<sup>1</sup> More than that, I hoped to derive from the logic of externalities the rationale for all government intervention in the workings of the economy — notions later echoed in works of Buchanan and Tulloch and the writings of Mancur Olsen. In the hotbed of discussion that constituted LSE the dissertation's ideas were very much a group product. We had made an effort to revive the famous Cambridge, LSE, Oxford seminar of the 1930s, and these subjects were discussed at the joint sessions, bringing in ideas from outside the LSE. Jan de v. Graaff played a leading part, he and I spending large amounts of time discussing one another's work on welfare economics.

In the course of that seminar's travels, incidentally, I also met Dennis Robertson, Joan Robinson, R. F. Kahn, J. R. Hicks, Lionel McKenzie and others whose names already were or were about to become legendary.

At LSE I gave two courses. Lionel Robbins invited me to lecture on economic dynamics, and my lecture notes for that course formed the basis for what was to be my first book, *Economic Dynamics*, which still survives. But in return for that plum I had to give a set of lectures on the

<sup>1</sup> I am not certain the term "prisoner's dilemma" had yet been invented; indeed, I seem to remember it was proposed by Professor A. Tucker at Princeton several years later. But the ideas of game theory had reached London and in discussions externalities were already being translated there into game theoretic terms.

American economy, a subject on which, most regrettably, I was dismally ignorant and which, I am sure, were an embarrassment to everyone, even if most educational to me. Much of the dynamics course, like so much else, followed Paul Samuelson's pathbreaking explorations which I merely translated into forms more accessible to students. People have sometimes been kind enough to suggest that my writing on the subject is rather clear. I have responded that I tried to write the book so clearly that even I could understand it — and that was, in all seriousness, the truth of the matter.

In both my lectures and the dissertation there were substantial sections on the relevant *dogmengeschichte* on which Lionel Robbins was enormously helpful. The graphic translation of the dynamics of the Ricardian model seemed to me an obvious interpretation of a system universally understood within the profession and generally agreed upon. I was amazed to find it singled out for special commendation in John Williamson's presidential address to the American Economic Association the year after publication, and to see it cited by writers on the history of thought many times thereafter. Since then it has become clearer that the substance of Ricardian economics is far from being agreed upon, as my former student Samuel Hollander has found in his debates with the neo-Keynesians of Cambridge and Italy and even with George Stigler and Paul Samuelson.

Our happy three years in London were drawing to an end. We had used the opportunity to travel about Europe a good deal. British exchange controls in that period of postwar austerity prevented us from using either my small earnings or my wife's comparable income, both in sterling, for the purpose. Fortunately, I also received support from the U.S. Government in U.S. currency, as an army veteran, and that permitted our first visit to Italy, which we have loved ever since. Of course, young graduate students did not expect many comforts while traveling or in their domestic arrangements so a small income served very adequately. There were severe shortages and strict rationing still continued. In London the foreign students and their spouses shared any packages received from abroad with their English friends. Often we would all pick up our six-week ration of one egg and one slice of bacon at the same time and get together to celebrate the feast. Heating was a great problem with tightly rationed coal of poor quality burnt in an open fireplace the main source of warmth in the home. One winter our water pipes froze for three months and produced a great indoor flood when they thawed and burst in the beautiful spring of 1947.

The dissertation was completed on schedule, and I had the most delightful oral examination on record over whiskeys and sodas at the Reform Club with my examiners, Marcus Flemming and Lionel Robbins who still considered my pursuit of the Ph.D. an American aberration. Professor Robbins knew we were determined to return to the United States, but he nevertheless made me a very generous offer at the LSE if I were prepared to remain. After I had refused with thanks most deeply felt, he recommended me with his characteristic kindness to Friedrich Lutz who was then visiting LSE from Princeton. Within weeks I had received and accepted an offer of an assistant professorship for the following academic year, and have remained at Princeton ever since.

Lifelong friends were acquired at the LSE — Lord and Lady Robbins and their children, Sir Arthur and Gladys Lewis to whom we had the pleasure of extending their first dinner invitation after they were married, Anne Bohm who brought sanity and order to the postgraduate program, Frank and Dorothy Hahn, and others as well. We have since made still more close friends in London where our visits (or theirs to the U.S.) have assumed the nature of reunions. In particular, Lord Robbins in his role as Chairman of the Royal Opera at Covent Garden and as Director of the National Gallery has over the years provided us with access to London's cultural activities such as few are privileged to enjoy.

### Princeton University (1949- )

At Princeton we were immediately welcomed into the community and the department. Richard Lester was chairman, and he quickly made us feel at home. Lionel Robbins had also written ahead about us to Jacob and Frances Viner with whom we remained on closest terms for the rest of their lives. I did not know Viner's terrifying reputation (which was belied completely by his natural but often deliberately concealed kindness). Consequently, I was foolish enough to disagree with him avidly and energetically whenever it seemed appropriate to the astonishment of many of our colleagues. Viner, who was obviously unused to such a response, was delighted. We spent many hours each week locked in debate. From time to time I needed information on economic history or on the history of ideas and was always very pleased when in response to literally any such question he would talk without

interruption for at least an hour providing an amazing stream of illuminating material from his bottomless stock of knowledge. He would also regularly, but apparently casually, make some unsupported theoretical statement which I was sure was quite incorrect and which he would then challenge me to disprove. Each time, after a considerable struggle on my part, the mathematics showed unambiguously that he was right. Years later a friend of his in another city recounted how Viner told him of a young friend on the faculty for whom he had set himself the task of presenting a new, paradoxical proposition in economic theory every week. What an education that was.

The following year Lester Chandler joined the Department, and from him I learned what little I understand about money and banking. Later we wrote a textbook together which, though it was unsuccessful, was an introduction to the pleasures of collaboration.

On arriving at Princeton we met two advanced graduate students, Martin Shubik and Harvey Leibenstein, with whom we have remained close friends. Within a year or two the department attracted an extraordinary group of undergraduate students including Richard Quandt, Otto Eckstein and Gary Becker with whom I wrote an article on Patinkin's dichotomy analysis and the pertinent materials in classical and neoclassical economics. We felt, and I still do, that Patinkin's discussions of the invalidity of the dichotomy between the real and monetary sectors of the economy, and of the problems caused by the assumption that supply and demand functions are homogeneous in prices alone, were brilliant pieces of work and constituted an extremely illuminating contribution. But at the same time some of the classical and neoclassical authors he accused of the resulting errors were, in my view, quite innocent of the charges. It should be added that the dispute was conducted just as such disputes should be, and that Professor Patinkin and I have become good friends and see one another both in Israel and the United States.

### Consulting Activities

In about 1953 Paul Lazarsfeld, the great sociologist then at Columbia University, asked me to work with him on one of his research projects. I quickly accepted the invitation, anxious to get to know more

about his path-breaking work, and, of course, because of the additional income which greatly relieved the assistant professor's traditional poverty. A key premise in much of Lazarsfeld's analysis was that while human behavior is stochastic, the relevant probabilities of the different options in tomorrow's behavior depend on the state of affairs today. Thus the probability that some individual will vote socialist in the next election will depend on whether he voted as a socialist or a Christian democrat in the preceding election. This immediately leads to the employment of the theory of Markov chains which yields exactly that sort of relationship. It also translates itself at once into a simultaneous system of stochastic difference equations whose coefficients are the relevant probabilities. It was this translation which had attracted Lazarsfeld to my writings, and the work I did for him was subsequently used in a substantial expansion of *Economic Dynamics* in its second edition. Later I also joined Lazarsfeld in two seminars, each lasting several weeks, one held in Switzerland and the other in the beautiful mountains north of Turin. There our families got to know one another.

Just after my work on the Lazarsfeld project came to an end I spent the spring and summer as a visiting professor at Berkeley. We traveled from Princeton to Berkeley by car with our two small children, making the country seem even more enormous than it is. At Berkeley we met many delightful people, notably Aaron Gordon, Robert Dorfman, (and for the second time) Harvey Leibenstein and their wives. I also met Joe S. Bain from whose work the theory of contestable markets would later derive so much, and Howard Ellis who was then approaching retirement.

Soon after our return to Princeton, where I had been promoted to full professor, I was introduced to another remarkable person, Wroe Alderson, who was then serving on an advisory committee to the Princeton economics department. Alderson was Quaker and a devoted advocate of their social goals such as the promotion of peace and elimination of poverty. He was senior partner of Alderson and Sessions, a management consulting firm in Philadelphia. After one evening together at a meeting he invited me to come to his office to see whether a consulting arrangement would suit us both. That arrangement lasted for nearly a decade. I would travel to Philadelphia about once a week and work on one or more of the many projects on which the busy company was engaged. In those years I got some sense of the way big business in the U.S. is conducted. I worked with large firms and small, among them major firms in chemicals, steel, food products and a variety

of other lines. I studied their policies on pricing, advertising, product line, location and all of the other areas of interest to economic analysis. The work made use of demand theory, mathematical programming, inventory theory and many of the other tools of formal economic analysis. It was a very valuable experience, teaching me how theoretical instruments can be applied flexibly to the complex and messy problems of reality, and, above all, suggesting how firms actually behave in reality. It also led me to revise many of my theoretical ideas, and two of my books originated from my work at the consulting firm. *Economic Theory and Operations Analysis* (1961) started off as a compendium of the analytical tools that had proved useful in application to business activities, with illustrations derived from experience with firms. In its later editions the book has gradually evolved more into exposition of economic theory with particular emphasis on very recent developments. The relationship between business experience and theory was still more direct in the case of *Business Behavior, Value and Growth* and its sales-maximization model. Several years of association with members of the managements of large firms finally forced me to recognize that there was a systematic difference between the way matters were viewed by them and by the standard economic models. Of the recommendations made to our clients, my impression was that about 60 percent were accepted and adopted. It became clear, eventually, that those which were rejected were not chosen fortuitously, but, rather, constituted a fairly predictable pattern. It ultimately dawned on me that virtually any proposal that promised to increase profits but did so by sacrificing sales volume was almost certain to be spurned. I began to modify my recommendations accordingly and, as I remember it, the acceptance rate rose substantially.

The natural reaction of a microtheorist to such experiences was a reexamination of the standard profit-maximization models which we had all been taught to employ. I found out, eventually, that other economists had, on the basis of interviews and other forms of observation, also reached the conclusion that in practice firms pursue objectives more diverse and complex than just maximization of profits. But in one important respect the consulting experience had taken me beyond that observation alone, for it had shown that other goals such as maximization of sales or growth of assets each had their own implications for business decisions and to each of these there seemed to correspond optimal values of the firm's decision variables. In other words, abandonment of the profit-maximization premise did not leave one with

no choice but chaos and indecision. Rather, it called for decisions which are different but equally determinate, and which can be analyzed using all of the traditional tools and methods — marginal analysis, mathematical programming, etc.

Some months of working on the formal analysis provided the model of sales maximization (later supplemented by a model of growth maximization) which constituted the basis of the little book *Business Behavior, Value and Growth* (1959). If this model does constitute a contribution I believe it consists not in the observation that management may have objectives other than profit, but in the demonstration that other objectives are perfectly consistent with fruitful theoretical analysis, as was later demonstrated so effectively by writers such as Oliver Williamson and Robbin Marris.

Let me emphasize that I never maintained and do not believe that all firms (or even all oligopoly firms) seek to maximize sales, or that they all share any other common and simple objective. I merely asserted and still believe that many firms have some objectives other than profit alone, and that for some which I encountered sales maximization is a reasonable approximation to their somewhat more complicated goals. Those goals are, in any event, rarely formulated expressly (except for purposes of public relations), they may change from time to time, and they are at most pursued only in a rough and ready manner. Moreover, I believe that in aggregative studies of industry behavior the differences between the predictions of profit and sales maximization are apt to be minor and unimportant. But for analysis of the behavior of individual firms I believe the distinction is vital, as the reactions of business clients to our recommendations made under the two premises suggests strongly.

Partisans of the profit maximization approach have suggested that the behavior patterns called for by the two models may be difficult to distinguish, and they have questioned the evidence presented in *Business Behavior, Value and Growth* which, admittedly, is all anecdotal. They have pointed out, rather cogently, that long-run profit maximization may require resistance to elimination of unprofitable sales in the short run. I certainly cannot prove the contrary. I can only suggest that the change in the nature of our recommendations constituted what amounts to a (very poorly) controlled experiment of a sort rarely possible for economists. Moreover, the resulting observations were generally supported by careful discussion with businesspersons with whom I had established rather close relationships and who had little reason to slant their answers.

After the sale of Alderson Associates when Wroe Alderson retired my consulting work continued through the firms Mathematica, and Consultants in Industry Economics (CIE), which I helped to found. This work has usually proved fruitful and stimulating to me in more general research. Microeconomists are peculiarly fortunate in this respect, for while academic consultants drawn from other disciplines generally contribute by application of the learning they have acquired through research and teaching, for them benefits rarely seem to flow in the other direction.

### The Performing Arts

Sheer misunderstanding led to my involvement in the economics of the performing arts. In about 1960, the Twentieth Century Fund and John D. Rockefeller III had decided that the time was auspicious for a systematic study of that subject. On inquiry they were told of an economist at Princeton who was knowledgeable about the arts as well as economics. The person who had steered them to me had, of course, confused my activities in painting and sculpture with knowledge of the finances and organization of opera, theaters, orchestras and dance companies. My father had instilled in me a great love of the performing arts but little knowledge of the economic side of these activities accompanied my wife's and my frequent attendance.

I agreed to a meeting on the subject, having first discussed the matter with a young colleague, William G. Bowen, with whom I had worked earlier (and who has since become President of Princeton University). Talking with the potential sponsors of the study I emphasized my ignorance of the subject and then, on the basis of my consulting experience, proceeded to indicate how I believed it should be analyzed dispassionately, as though one were dealing with the economics of the most banal of commodities rather than one which commands widespread expressions of adulation (if hardly universal attendance).

As it happened, Bowen and I were then both heavily committed to other projects. When it turned out that the potential sponsors had, for better or worse, decided that we were the right persons to conduct the study, we found ourselves resisting what was to prove one of the most exciting projects we had ever undertaken — so poor can foresight be!

The research turned out to be a major undertaking. It lasted somewhat more than three years, acquired and analyzed data from several hundred organizations, distributed questionnaires to about 150,000 audience members at more than a hundred performances in dozens of cities. A small army of students and other investigators was trained and sent out to collect the materials using questionnaires which had been painstakingly designed and pretested. All of this was planned meticulously by Bowen and the work was organized and supervised by my wife, with whom all my subsequent work on the economics of the arts has been carried out in full partnership.

The overall design of the study and the planning of the empirical work is all to be credited to my coauthor. His skill in dealing with messy data made it possible to determine many fundamental attributes of the activities in question. For example, he was able to show systematically that the composition of audiences in terms of education, income, age and sex varied only minusculely from one art form to another or one city to another. Audiences in London were, essentially, no different from those in Houston, Texas — all highly educated, and well-to-do relative to the population as a whole. Generally no more than two or three percent of the audience was made up of blue collar workers; opera with seven percent of the audience derived from this economic group, constituting the only exception. Many fascinating observations peripheral to our central topic also emerged from Bowen's calculations. For example, he was able to show that the proportion of women among the performance in an orchestra at that time was almost perfectly inversely correlated with the income of the orchestra — a very tangible index of sex discrimination!

In the early 1960s collection of economic data on the performing arts in the U.S. was no routine matter. It often took my wife to back offices, basements and lofts, where figures sometimes had to be pieced together from scraps of paper that constituted the only records that some group had kept. Since that time much more such information has been collected and some of it published on a regular basis. Indeed, we were asked to design and, for several years, collect some of the figures that now appear regularly in the official U.S. government publications.

My own part in the study involved a role in the determination of the overall objectives and the general research design. I also wrote most of the final manuscript. However, my one contribution I consider to be significant is the cost disease model, which has since been used to help explain the behavior of the cost of education, the budgetary problems of

cities, etc. The basic point is that the *live* performing arts (in contradistinction to those employing the mass media such as film and television) are extraordinarily unadaptable to productivity-increasing technical change. A Boccherini string quartet written in the eighteenth century which takes half an hour to perform, required two-person hours of performance time then and requires exactly the same amount of time today. Meanwhile, in much of the remainder of the economy productivity has been increasing virtually without interruption in a manner that compounds and accumulates. If wages in the arts do not rise far more slowly than those in the remainder of the economy (and the evidence indicates that they do not) this means that cost per performance in the arts must rise steadily at a rate faster than costs in the remainder of the economy, roughly reflecting the difference in the productivity growth that characterizes the two sectors. Over the years this can add up to an enormous differential. I estimate roughly that a typical manufactured good which cost about the same as attendance at a performance in 1800 now costs only about one twentieth as much! This means also that the funds supplied to the arts by government or private philanthropy must generally increase year in, year out, at a rate exceeding the economy's rate of inflation if artistic activity is not to be forced to retrench. Constancy of *real* contributions is simply not enough for the purpose. This, then, is what has since come to be called the cost disease of the arts (I am delighted that it is also sometime called "Baumol's disease"). If the hypothesis is correct (and there is a good deal of evidence consistent with it)<sup>2</sup> it helps to explain many economic phenomena outside the arts, such as the shift of the labor forces in a number of countries out of manufacturing and toward the services, the increasing use of disposable products to avoid repair, the rising relative cost of medical care and education, etc.

Because such incidents are sometimes considered noteworthy it may be worth reporting that the cost disease model entered my consciousness quite suddenly and unexpectedly, though it had no doubt lurked in my subconscious for some time before. One night I awoke

<sup>2</sup> The noteworthy exception seems always to have been a period of rapid inflation. Data for the period after the U.S. Civil War, after World War I and the inflation of the 1970s indicate that in such periods growth in cost per performance has not stayed ahead of the general price level. Presumably groups are not able to raise money quickly enough, and are therefore forced to retrench by reducing rehearsal time, simplifying costumes and scenery, etc. If this explanation is correct, it follows that the cost disease is not really in remission in such periods, but that its normal consequences are suppressed and emerge in a different form — that of reduction of standards and cutting of corners.

from a deep sleep at about 3 a.m. with the entire model clearly in my mind. I left the bedroom, went to the next room, jotted down a few notes, and immediately went back to sleep. The next morning I was able to write the idea up systematically. Later I will recount another such incident which occurred about 15 years afterwards.

Our study of the arts was sponsored by a private foundation, the Twentieth Century Fund, which generously paid the considerable costs of this project. Moreover, the Fund was careful to avoid interference of any sort with the nature of the research or the contents of the resulting book. Indeed, it withstood pressure from a representative of some of the larger performing organizations to suppress or weaken some of the materials showing the high incomes of the members of the audience, a piece of information which, it was feared, would make it more difficult to obtain financial support from the government and other sources. Happily, these fears proved to be unjustified.

As a matter of fact, it is now widely believed that the opposite occurred — that because of the accident of timing (or, perhaps it was no accident but a matter of good judgment of its sponsors) the book is now said to have played a role in launching substantial government support in the United States. It appeared at the same time as another report sponsored by the Rockefeller Brothers Fund (*The Performing Arts: Problems and Prospects*, McGraw Hill, New York, 1965), in which Bowen and I had a peripheral part. The Rockefeller report was prepared by a committee composed to a considerable extent of businessmen whose strong statement of approval of funding for the arts contributed political respectability to the attempt to induce the United States to embark upon the sort of public sponsorship so traditional in Europe. The simultaneous appearance of our own book, which for the first time provided systematic data and some degree of analysis on the subject, apparently contributed ammunition to the campaign for public support of the arts.

The book was launched with a burst of publicity orchestrated by the Twentieth Century Fund. There were front page stories in the *New York Times* and the *Washington Post*. Newspapers throughout the world carried substantial reports. We were most amused by the story in *Pravda* which reported that two respectable economists from Princeton University had just published a book showing how capitalism destroys the arts!



### Princeton and New York Universities

The economics department at Princeton was an extraordinarily harmonious group, indeed. I believe virtually all of my colleagues had a deep affection for one another. We were sorry when Friedrich Lutz returned to Europe a few years after we arrived. In addition to members such as Viner, Oskar Morgenstern, Richard Lester, and Lester Chandler, the Department was fortunate to acquire W. Arthur Lewis and Fritz Machlup, both of whom were to become close friends. The Lewises had their home a few doors from ours, and our children grew up together. His lovely and charming wife also proved to be my most talented sculpture student, and has had several successful exhibitions of her own. Sir Arthur's general wisdom, his analytic intuition and his meticulous historical scholarship have always constituted standards difficult for the rest of us to emulate. Fritz Machlup, whose unexpected death earlier this year has left a major gap in our lives, was an incredible generator of ideas and research undertakings which he pursued with amazing determination and energy. At age 77 he undertook a research program which was projected to yield a work in ten volumes. Three years later he had completed the first three of those books, and work on the others was well on its way.

He and Oskar Morgenstern reached retirement age virtually simultaneously. On attaining their emeritus status at Princeton they were immediately offered and accepted full term appointments at New York University, where each of them continued to teach with energy and devotion for the remainder of his life.

At just about the time Machlup and Morgenstern left for N.Y.U. I had begun to think about my position. By then I had been at Princeton for nearly a quarter of a century. Our children had grown and left home. It seemed high time to make some change in our arrangements. Yet leaving Princeton altogether was virtually unthinkable. Half our lives had been spent there and it contained most of our close friends, both outside the department and within it.

The solution to our dilemma was an invitation to N.Y.U., first, to come as a visitor for a year. Aside from visiting appointments at the Stockholm School of Economics and Berkeley this was our first protracted period away from Princeton since 1949, and it was a return to the city of our childhood. The department in New York proved most hospitable and a pleasant place in which to work. The students were far more heterogeneous than those at Princeton, both in ethnic background

and in quality of their earlier training. The best of them were of outstanding quality, and they constituted a fascinating and enthusiastic group.

Consequently, when the next year I was offered a permanent appointment on a half-time basis, and received the consent of Princeton to the arrangement, I agreed readily. As it turned out, it was the beginning of what, in my own view, seems my most creative period.

It actually began with the completion of an exclusively Princeton enterprise; two volumes on environmental economics written jointly with my (then) Princeton colleague, Wallace Oates. One volume was almost entirely devoted to theory while the second sought to assemble the available empirical materials, adding some evidence acquired by ourselves.

Among the results provided by the theoretical volume two will be cited as illustrations. The first deals with the victims of externalities. It has, of course, been known, at least since the work of Pigou, that (with some restrictions) optimal expenditure on reduction of pollution emissions (or other detrimental externalities) requires a charge or tax upon the polluter equal to the marginal social damage of his emissions. This 'polluter pays' arrangement forces the polluter to bear the full social cost of his emissions, and therefore, to undertake any preventative measure whose (incremental) cost is less than the value of the damage thereby avoided. But what about the victims? It would seem that simple justice calls for some sort of compensation to those whose health or even whose cost of living is affected by pollution. Indeed, this intuitive judgment is readily confirmed formally with the aid of the mathematical theory of fairness provided in the work of Duncan Foley, David Schmeidler, Serge-Christophe Kolm and others. Yet it is proved in our theoretical volume that, unless it can be provided in a way whose incentive effects are zero (i.e., as a "lump sum" payment), *any* compensation to the victims of externalities, however small, is incompatible with Pareto optimality in resource allocation in the economy. There is a simple theoretical explanation of this result. In the presence of detrimental externalities there is a Pareto optimal level of use of resources *by the victims* to protect themselves from their effects. For example, it may be appropriate to insulate or air condition homes and workplaces, or it may be desirable to move them away from the source of pollution thereby, perhaps, increasing transportation costs in the future. But if compensation is based on the amount of damage suffered by the victims, such compensation payments will induce them to spend less on self protection against that damage. In effect, such compensation

payments reduce the marginal net yield of outlays on self-protection. Here we then have a clear example of a conflict (or, rather, a trade-off) between Pareto optimality and fairness.

A second illustrative result of our theoretical analysis relates to the choice between a tax upon the generator of externalities and a subsidy to induce him to reduce his emissions. Common sense suggests that at appropriate tax and subsidy rates the effects of the two will be the same — the donkey can be moved either with the carrot or the stick. But while this conclusion (which had often been repeated in the environmental literature) has an element of truth, it turns out that subsidies have a second consequence which is likely more than to offset its pollution decreasing effect. It is true that in the short run either the tax or the subsidy will induce polluting firms to emit less. However, at least in the case of perfect competition, in the long run, the tax on emissions may *not* reduce the emissions of any one polluting *firm*, yet it will reduce the emissions of the *industry* by encouraging the exit of polluters. On the other hand, *the subsidy is likely to increase the industry's emissions* by encouraging the entry of polluters. These results are certainly true when there is a fixed proportion between output and quantity of emission. A fixed Pigouvian tax,  $t$ , per unit of emissions then contributes to the firm's total cost the amount  $tky$ , where  $y$  is output and  $k$  is the emissions-output ratio. The resulting addition to average cost is  $tky/y = tk = \text{constant}$ . Therefore, the Pigouvian tax simply causes a uniform upward shift in the firm's average cost curve, and so does not affect the location of its minimum point, i.e., its profit maximizing output or emissions level. Yet we do know from a standard supply-demand diagram that with curves of the usual shape a tax must shift the supply curve upward and therefore must reduce industry output and, hence, its emissions, and the opposite must be true under a subsidy.

The investigation of the empirical data also produced a number of what, to us at least, were surprises. We had thought that the explosion in the world's population and in industrial activity would show that there was fairly universal growth in the rate of environmental damage, but the results were far more mixed. In some cases such as lead in the atmosphere and the generation of solid wastes our conjecture did indeed turn out to be true or, at least, to be supported by the evidence. But postwar emissions control efforts had substantial beneficial effects on air and water quality with, e.g., enormous decreases in the sulphur and particulate content of the atmosphere in major cities of the U.S. and the United Kingdom. In other cases, e.g., the concentration of pollu-

tants in Lake Superior in the United States and the oxygen content of the rivers surrounding Manhattan, matters had been improving long before that. In several cases such as the depletion of oxygen in the Baltic and the concentration of mercury in tuna it transpired that there was good reason to suspect that the causes were natural (e.g., a secular decline in rainfall in the sources of water for the Baltic). We were most surprised to find (after an extensive search that took us to the conservators' quarters at the Louvre, the Metropolitan Museum in New York and St. Paul's Cathedral in London) that there seems to be no conclusive evidence supporting the view that the deterioration of ancient sculpture and of stone buildings is accelerating or that it is attributable to increasing emissions of pollutants. We found many newspaper stories claiming categorically that this was so, but careful review of the scientific evidence simply forced us to accept the verdict 'not proven'.

Perhaps partly in consequence of this work I was elected to the presidency of the recently formed Association of Environmental and Resource Economists (AERE), an organization which has since grown and expanded its activities.

### The Theory of Contestable Markets

Soon after the two books on the environment<sup>3</sup> made their appearance, I found myself embarked almost accidentally on what I consider to be the most fruitful piece of research in which I have ever participated. The word 'participated' is used advisedly, since the results stemmed from the work of at least a half dozen persons beside myself — my coauthors, John Panzar, then of Bell Laboratories and Robert Willig of Princeton University, Elizabeth Bailey, formerly Vice Chairman of the Civil Aeronautics Board and now at Carnegie Mellon University, Dietrich Fischer, Thijs ten Raa, then at New York University, and Gerald Faulhaber at Bell Laboratories. Obviously there were three centers of activity: N.Y.U., Princeton and Bell Laboratories.

<sup>3</sup> WILLIAM J. BAUMOL and WALLACE OATES, *The Theory of Environmental Policy*, Englewood Cliffs, N. J.: Prentice Hall, 1975 and BAUMOL, OATES and S. A. BLACKMAN, *Economics and the Quality of Life*, Englewood Cliffs, N. J.: Prentice Hall, 1979.

My systematic work on the subject began with a project I had undertaken under the sponsorship of the Division of Information Science and Technology of the National Science Foundation in which an incidental part of the task was to provide a nontechnical document discussing the rationale for and principles for the determination of the proper amount of government financial support for the dissemination of scientific and technical information in general and of scientific journals in particular. Assuming that a large part of the argument would rest on the sorts of market failure associated with public goods and scale economies, I set about what I expected to be the tedious task of redescribing a set of elementary and straightforward principles. The public goods portion of the discussion turned out just as expected — exactly as it is so well described in the literature. However, the theory of scale economies and the associated phenomenon of natural monopoly seemed to resist simple explanation. Each time I attempted a description of the logic of some obvious proposition it seemed to acquire complexity and turned out not to be quite correct. For example, in the case of a publisher who provided a half dozen journals rather than only one, scale economies did not seem to account for the firm's multiproduct character. Why was the enterprise a multiproduct firm, and what would society lose by breaking up the publishing firm into six separate publishers of single journals? It was considerations such as this which later led Panzar and Willig to formulate their concept of economies of scope — the savings which a firm may (or may not) enjoy from simultaneous provision of a multiplicity of products — and which had previously led me to formulate the more technical concept of trans-ray convexity,<sup>4</sup> a formal criterion of continuous complementarity in the production of different goods as output proportions change. At the same time it became clear that natural monopoly which is defined to mean that production of the industry's vector of outputs is cheaper when carried out by a single firm rather than by any multiplicity of firms, is not quite the same thing as scale economies. Indeed, it was proved eventually that scale economies throughout the relevant region of output space are neither necessary nor sufficient for natural monopoly.

Meanwhile, Gerald Faulhaber, who had been sent by Bell Laboratories to carry out his graduate work at Princeton, had quite indepen-

dently, starting off from some earlier work of mine, begun research on the theory of natural monopoly. He found implications which had escaped me completely at the time the original work had been prepared. When Faulhaber showed me some of his work and asked me to supervise his Ph.D. dissertation, I was astonished at the degree of overlap with my own work at N.Y.U. Indeed, using a single-product game theoretic approach, his work was at that point in many ways ahead of mine.

The developments were sufficiently seductive to attract the attention of others, notably Panzar and Willig at Bell Laboratories, Dietrich Fischer at N.Y.U., and Elizabeth Bailey who then held positions both at Bell Laboratories and N.Y.U. and who quickly took on the task of liaison in addition to the valuable contributions she directly provided herself.

Faulhaber had already proved for the single-product case that the concept underlying and, indeed, appropriately defining natural monopoly is a mathematical relationship called subadditivity of total cost. Specifically (for the multiproduct case) let  $y^I$  represent the vector of outputs of the industry and let  $y^i$  be any vector of outputs assigned to a hypothetical firm,  $i$ , in any partition of the industry's outputs, so that  $\sum y^i = y^I$ . If  $C(y)$  is the total cost function of a firm in the industry, then that cost function is strictly subadditive if  $C(y^I) < \sum C(y^i)$  for each and every set of  $y^i$  summing to  $y^I$ . In other words, the cost function is subadditive at industry output vector  $y^I$  if it is cheaper for  $y^I$  to be produced by a monopoly than by any larger number of firms.

Faulhaber had also proved in the single product case, first, that economies of scale are sufficient to guarantee that the firm's average costs will decrease, second, that decreasing average costs are sufficient to guarantee subadditivity, and, third, that the converse is untrue — that subadditivity is no guarantee of decreasing average cost. In other words, in the single-product case he had shown that an industry could be a natural monopoly even if it did not exhibit scale economies and its average costs were not declining.

Soon Panzar and Willig produced several sets of necessary conditions for subadditivity in the multiproduct case, and I provided a set of conditions sufficient for multiproduct natural monopoly, conditions which have since been used rather widely in empirical studies of cost functions and their implications for the structure of an industry.

Panzar and Willig, working together, proposed a concept which Dr. Bailey and I had also put forth on the same day — the concept of

<sup>4</sup> For a formal definition see BAUMOL, PANZAR and WILLIG, *Contestable Markets and the Theory of Industry Structure*, San Diego, Ca: Harcourt Brace Jovanovich, 1982, pp. 79-81.

sustainability of prices. A vector of prices charged by a multiproduct monopolist is said to be *sustainable against entry* if a) it provides revenue to the monopoly at least sufficient to permit it to cover its costs, and b) there exists no other vector of outputs which an entrant can sell at those prices and which will permit the entrant to operate without loss. In other words, sustainable prices permit the incumbent to prevent entry without recourse to retaliatory measures or strategic responses. This concept was formulated as a first step toward transformation of the natural monopoly concept from one which was normative (when is natural monopoly the most efficient industry structure?) to something that was more or less behavioral (when will natural monopoly be immune from entry?).

Faulhaber was able to prove that where both economies and diseconomies of scale are present (at different ranges of outputs), even if a firm is a natural monopoly, no sustainable prices may exist for it. He did this with the aid of a remarkable numerical counterexample: Consider three communities for which electricity generating facilities of given capacity are to be built. Suppose any one of the communities' needs can be met at a cost of \$ 12 million, that any two of them can be served simultaneously for \$ 19 million while all three of them can be served by a single plant for \$ 30 million. Costs here are clearly subadditive since provision of the output by three separate plants costs  $3 \times \$ 12 \text{ million} = \$ 36 \text{ million}$ , while two-plant production costs  $\$ 12 + \$ 19 \text{ million} = 31 \text{ million}$ . Both other options are, therefore, more costly than the \$ 30 million outlay required for single-firm production. Yet, here, a single generating firm can find no prices which are sustainable. For if it were, for example, to propose to charge each community \$ 10 million in order just to cover its \$ 30 million cost, an entrant could offer to supply only two of the three communities at a cost of say \$ 9.7 million each, thus more than covering the \$ 19 million cost of the plant required for the purpose. Faulhaber's example thus showed that in some cases freedom of entry can prevent cost minimization. But the bulk of our analysis was later to argue strongly for ease of entry.

A little later I provided a complementary theorem which rested on a result first derived by Frank Ramsey in 1927. As we know, where average costs are declining (and in a variety of related multiproduct cases), a firm which prices its products at their marginal costs must suffer losses. The question to which Ramsey addressed himself is if marginal cost prices are therefore precluded, what deviations of prices from marginal costs are required for the second-best allocation of resources, that is, for Pareto optimality under the constraint that supplying firms just cover their total

costs? Ramsey was able to derive a formula for these second-best prices which was ignored by much of the economic literature for several decades but had by the 1970s received a great deal of attention. I was then able to prove (though the result took me completely by surprise) that a monopolist who decided to adopt Pareto optimal Ramsey prices would (under a set of rather reasonable assumptions) find those prices to be sustainable. In other words, such a commendable pricing decision would reward the monopolist by granting him immunity from entry.

I had arrived at this result in rather curious circumstances. My wife and I were attending a fund raising performance at one of New York's experimental theaters, and we were waiting in line to get in surrounded by persons in bizarre dress and make-up when, according to my wife, my face took on a rather distracted look. I told her that a theorem which hardly seemed plausible to me had come to me from nowhere, along with what seemed to me to be its entire proof.

Indeed, it transpired that the rigorous proof that emerged eventually did follow the outline that came to me suddenly in that theater lobby, but it took weeks of hard labor by (a skeptical) Willig, Bailey (and myself) before it could be put into satisfactory form.

At about this time a small group of U.S. economists was sent by the National Science Foundation to attend a conference in Leningrad, and Willig and I were among them. We sat up all night in the airplane discussing how our analysis of natural monopoly could be extended to other market forms and constitute the basis for a theory of the determination of industry structure. The vision had been mine, but Willig was to contribute the key step to its realization. First, on our return, Fischer and I produced a paper showing how one can calculate the cost minimizing structure of an industry, thereby determining whether it is or is not, say, a natural monopoly. For example, if it transpires that the industry's output vector can be produced most cheaply by say, four firms, one can say it is a "natural oligopoly", and we showed under what circumstances this will be true. Similarly, we showed under what cost conditions the industry will be "naturally perfectly competitive", etc.

Then Willig formulated the concept of what we were to call a perfectly contestable market — a market in which an entrant has access to all production techniques available to incumbents, in which the entrant is not prohibited from wooing the incumbent's customers, and in which entry decisions can be reversed without cost — that is, a market from which entrants can withdraw without loss of any of their

investments. An example of an approximation to such a market is an airline route. If company A opens up for business on the route from New York to Los Angeles and business proves disappointing, he can simply withdraw and move them to another more promising route.

Where exit and entry are so easy and exit so costless incumbents are completely vulnerable to the threat posed by *potential* entry. It is possible to prove that, as a consequence, in a perfectly contestable market a) no firm can earn any monopoly profit in the long run, b) industry structure must always be efficient, i.e., the industry will tend to be composed of exactly the number of firms that can produce its output at minimum cost and c) if two or more firms supply a given product to a market, in the long run the price of that product must equal its marginal cost.

Markets may be perfectly contestable even if they are characterized by scale economies in production and even if they contain only a small number of firms (even only a single firm). The theory thus generalizes considerably the concept of perfect competition, showing how ease of entry and exit and the accompanying threat by potential entrants can elicit good performance even in industries with small numbers of firms.

It must be emphasized that our purpose was not apologetics. We do not believe that most industries are perfectly contestable or even nearly so. We do believe, however, that some industries with small numbers of firms *are* highly contestable, and that in those cases government interference with the market mechanism is difficult to justify. In other cases the contestability of the market can be increased by public policy and in those cases this will sometimes prove to be the most effective means to serve the public interest.

It should be added also that we are well aware of the heavy debt contestability analysis owes to earlier writing, and have tried to suggest some of its sources in our publications. But as Viner taught me long ago, one can never hope to achieve completeness in such an undertaking.

In my presidential address to the American Economic Association in December of 1981 I sought to provide a general introduction to the theory of contestable markets. This was followed several months later by the publication of our book with its lengthy analysis of multiproduct firms and industries and its examination of the market forces that determine the structure of an industry — whether it will emerge as an oligopoly, a monopoly or something else. As is to be expected, the analysis has generated controversy, raising legitimate questions many of

which are still far from being settled. It has also led to a variety of research undertakings by others, both empirical and theoretical, and that has, of course, been most gratifying.

### Toward Further Work

Happily, the area in which I work has experienced relatively rapid growth in the past few decades, and so my stock of teaching capital has undergone considerable obsolescence. Many of my former students and their contemporaries are now better equipped than I to teach the various courses in mathematical economics, virtually all of which I had inaugurated some thirty years earlier.

I flatter myself that this is not quite true in my research. In the two years since the appearance of the contestability book I have embarked on several other projects. I am working on applications of Duncan Foley's fairness concept to analyze the equitability of different rationing procedures, of peak-off peak pricing and of Pigouvian taxes on externalities. I am working on the theory of productivity growth and the feedback relationship between such growth and expenditures on research and development by private industry. I am also considering a study of the theory of nationalized industry — of the circumstances under which operation of a firm by the public sector may be superior in terms of the general welfare to operation by private enterprise, regulated or unregulated by government. Here it should be noted that analyses such as the theory of public goods are less pertinent than may at first appear to be the case since a public good can be produced by private firms if it is *financed* by government. If nationalized firms lack the incentives for efficiency provided by the market mechanism to private firms why, then, should government enterprise ever be preferred over private? I am working on a model which will, with a bit of luck, provide some answers and, perhaps, some additional insights.

In short, as yet there is no conclusion to my story — all I can offer is a status report on a continuing stream of research...

*New York and Princeton  
Universities*

WILLIAM J. BAUMOL