Preface

to *Increasing Returns and Path Dependence in the Economy*,

W. Brian Arthur, University of Michigan Press, 1994

When the University of Michigan Press approached me to bring out a book of collected papers on increasing returns in economics I was surprised. I had thought that only older researchers, venerable and near retirement, issued collected works. But Timur Kuran, my editor, and Colin Day, the Press's director, argued that although the papers collected here have been receiving much attention lately, several of them have appeared in obscure places and are not easy to track down. In book form they would be accessible. Moreover, if they were brought together, a wider picture might emerge from the mosaic they create than from the individual pieces. This sounded like sufficient rationalization to me, and I accepted their invitation with gusto.

Ideas that invoke some form of increasing returns are now acceptable in economics—indeed they have become highly fashionable. But this was not always so. As recently as the mid-1980s, many economists still regarded increasing returns with skepticism. In March 1987 I went to my old university, Berkeley, to have lunch with two of its most respected economists. What was I working on? Increasing returns. "Well, we know that increasing returns don't exist," said one. "Besides, if they do," said the other, "we couldn't allow them. Otherwise every two-bit industry in the country would be looking for a hand-out." I was surprised by these comments. Increasing returns *did* exist in the real economy, I believed. And while they might have unwelcome implications, that seemed no reason to ignore them.

Since then much has changed. The whole decade of the 1980s in fact saw an intense burst of activity in increasing returns economics. Nonconvexities and positive feedback mechanisms are now central to modern theorizing in international trade theory, the economics of technology, industrial organization, macro-economics, regional economics, growth theory, economic development, and political economy.

Of course this turnabout did not happen in just a decade; it has been coming for a long time. In a sense, ideas that made use of increasing returns have *always* been part of the literature in economics. But in the past they were only partially articulated and were difficult to bring under mathematical control. And they tended to have disturbing implications. As a result many in our profession chose to

disregard or dismiss them. This distaste reached its peak in the early 1970s with the broad acceptance in economics that all properly specified economic problems should show a unique equilibrium solution. I was a graduate student about this time, and all results in economics were served to us with the incantation that they were true, "providing there is sufficient convexity—that is, diminishing returns on the margin." I was curious about what might happen when there were *increasing* returns on the margin, but none of my professors seemed interested in the question or willing to answer it. Examples with increasing returns and non-convexities were of course mentioned from time to time. But in the main they were treated like the pathological specimens in labeled jars that used to be paraded to medical students—anomalies, freaks, malformations that were rare, but that nevertheless could serve as object lessons against interference in the natural workings of the economy.

Part of the change, part of the very acceptance of the place of increasing returns, came out of the more formal part of economics itself. General-equilibrium and game theorists have known for many years that even under the most benign assumptions, multiple equilibria and indeterminate solutions occur naturally in the problems they deal with. Hence they found little difficulty accepting multiple equilibria when they arose from increasing returns. International trade theory, a less formal part of economics, needed to explain the peculiarity of intra-industry trade—France selling electronics to Germany and Germany selling electronics to France—and to deal more realistically with trade in manufactured goods. This forced it to include the possibility of increasing returns in production. Just as important as the broader acceptance of nonconvexities in trade theory and mathematical economics has been the development of new methods that deal analytically with the market imperfections and stochastic dynamics that arise in increasing returns problems. Economists can now explore into the terrain of increasing returns with much better equipment than they could ten or twenty years ago. As they do this they are rediscovering outposts from earlier expeditions: not just Adam Smith's familiar writings on specialization, Myrdal's notion of cumulative causation, Kaldor's work on mechanisms of regional disparity, Rosenstein-Rodan's Big Push; but Marshall's hopes for an organic economics, and the ideas of Frank Graham, Piero Sraffa, Allyn Young, Tord Palander, Magoroh Maruyama and others more obscure.

All the essays brought together in this book were written between 1982 and 1992. In selecting papers to include I have dropped ones that were highly mathematical and ones that repeated arguments made in the papers here. Even so, there remains some overlap in the material included. Many of the papers have undergone several rewrites, each time becoming more compressed, more technical, and less discursive as they neared publication. In several cases I have deliberately selected an early version, in the hope of preserving some of the liveliness that later got edited out. And I have arranged the papers not by subject but more or less in the order in which they came to be written, with some exceptions to help the exposition along.

There are of course other approaches to increasing returns, notably the imperfect-competition, static approach prominent in international trade theory, pioneered by Elhanan Helpman, Paul Krugman

and others, and the deterministic-dynamic approach of Paul Romer and others who have explored endogenous growth powered by increasing returns mechanisms. These other approaches are important; but they are not covered in the papers collected here.

The papers in this collection reflect two convictions I have held since I started work in this area. The first is that increasing returns problems tend to show common properties and raise similar difficulties and issues wherever they occur in economics. The second is that the key obstacle to an increasing returns economics has been the "selection problem"—determining how an equilibrium comes to be selected over time when there are multiple equilibria to choose from. Thus the papers here explore these common properties—common themes—of increasing returns in depth. And several of them develop methods, mostly probabilistic, to solve the crucial problem of equilibrium selection.

*   *   *   *

As part of bringing together these essays in book form, I have been asked by the editors to give an account of how I came upon the ideas they contain.

My serious involvement with increasing returns did not begin until 1979. Before that, however, a couple of things had hinted in the right direction. I had studied electrical engineering as an undergraduate, so that the notion of positive feedback was familiar to me; though I remained vague about its consequences. And I had studied graduate-level economics at Berkeley, where I had become fascinated with the question of economic development; so that I was familiar with mechanisms that involved cumulative causation and self-reinforcement in Third World economies. As a result I remained curious about positive-feedbacks and increasing returns in the economy; but I could not see how they might be incorporated into theory in a general and rigorous way.

This changed in 1979. I was working in the theory group of the International Institute for Applied Systems Analysis (IIASA) in Vienna and in April and May of that year got leave to spend eight weeks in Hawaii at the East-West Center. On my way through California, I picked up a copy of Horace Freeland Judson's *The Eighth Day of Creation*, a beautifully written, seven-hundred page history of the discovery of the structure of DNA, the deciphering of the genetic code, and the discovery of the structure of the hemoglobin molecule. I became fascinated with Judson's book and in the next couple of weeks absorbed it in detail. This led to my reading in Hawaii whatever I could find in molecular biology and enzyme reactions. Among the books I got my hands on was Jacques Monod's *Chance and Necessity*, an insightful essay on the interplay between determinism and historical accident which was inspired by his discoveries of autocatalytic reactions that could go in more than one direction. Back in Vienna, in the fall of 1979, I followed some of these ideas from biochemistry and molecular biology into the domain of physics. A colleague, Mark Cantley, told me of the work the Brussels group had been doing on enzyme reactions, and lent me Ilya Prigogine's essay "Order through Fluctuation: Self-Organization and Social System." I began to learn thermodynamics so that I could

study the work of Glansdorff, Nicolis, Prigogine, and others. At this time I also studied in detail the work of the German physicist Herman Haken.

That this body of theory represented a different point of view on science was clear to me at once. In this work outcomes were not predictable, problems might have more than one solution, and chance events might determine the future rather than be averaged away. The key to this work, I realized, lay not in the domain of the science it dealing with, whether laser theory, or thermodynamics, or enzyme kinetics. It lay in the fact that these were processes driven by some form of self-reinforcement, or positive feedback, or cumulative causation—processes, in economics terms, that were driven by nonconvexities. Here was a framework that could handle increasing returns.

A great deal of my approach to increasing returns problems and to economics fell into place in a few weeks in October and November 1979. The problems in economics that interested me, I realized, involved competition among objects whose "market success" was cumulative or self-reinforcing. I discovered that wherever I found such problems, they tended to have similar properties. There was typically more than one long-run equilibrium outcome. The one arrived at was not predictable in advance; it tended to get locked in; it was not necessarily the most efficient; and its "selection" tended to be subject to historical events. If the problem was symmetrical in formulation, the outcome was typically asymmetrical.

In individual problems, some of these properties (especially the possibility of non-efficiency especially) had been noticed before. But there did not seem to be an awareness that they were generic to increasing returns problems and that they might form a framework for discussion and dissection of such problems. Further, it seemed that these properties had counterparts in condensed-matter physics. What I was calling multiple equilibria, non-predictability, lock-in, inefficiency, historical dependence, and asymmetry, physicists were calling multiple meta-stable states, non-predictability, phase or mode locking, high-energy ground states, non-ergodicity, and symmetry breaking.

I became convinced that the key obstacle for economics in dealing with increasing returns was the indeterminacy introduced by the possibility of multiple equilibria. A statement that "several equilibria are possible" did not seem acceptable to economists. Missing was a means to determine how a particular solution might be arrived at. What was needed therefore was a method to handle the question of how one equilibrium, one solution, one structure, of the several possible came to be "selected" in an increasing returns problem. One possibility—popular at the time in game theory—would have been to add axioms that would settle the "selection problem." But this seemed artificial. Selection should not be predetermined I believed; in most problems it would happen naturally over time, often by historical accident. Thus the approach I sought needed to allow the possibility that random events, magnified by the inherent positive feedbacks or reinforcing mechanisms, might select the outcome probabilistically in increasing returns problems. Equilibrium selection, I believed, should be modeled by using non-linear stochastic processes.

For a time master equations from probability theory seemed promising. But they were unnecessarily complicated and not very general—for example, they could not easily treat processes with growth. It became clear I needed to tailor my own methods.

In 1980 I experimented with various stochastic formulations of my economics problems with increasing returns, with mixed success. In the summer of 1981 I invited Joel Cohen, the mathematical biologist from Rockefeller University, to IIASA, and told him of my difficulties trying to find a suitable probabilistic framework to embed my problems in. Joel put me on to the Polya process, a path-dependent process in probability theory. He had in fact recently written a classic piece on the Polya process as a metaphor for how historical accidents could determine future structures. I was taken with the cleanness of the Polya framework, but realized that my problem did not exactly fit its specifications.

In fact, the Polya process turned out to be a very special case. It provided a growth process where units were added one at a time randomly into different categories with probabilities identical to their current proportions. What I needed to work with in my market-build-up problems was growth processes where probabilities of addition to the categories could be *an arbitrary function* of their current proportions (or market share). I believed that such "allocation processes" would converge to fixed points of this probability function, and convinced myself of the correctness of this conjecture using Fokker-Planck techniques. I asked a number of professional probability theorists for help in providing a rigorous proof. No one could oblige. I had mentioned the problem several times to my office mate at IIASA, the Soviet probability theorist Yuri Ermoliev. One day Yuri asked me to be show him the formulation one more time; he thought he might have a possible idea that might point toward a proof. Ermoliev's idea was to reduce the dynamics of the process to a format that was well-understood in probability theory—that of the Robbins-Monroe stochastic approximation. It looked as if it would work. Ermoliev farmed the task of nailing the proof out to his protege, Yuri Kaniovski, and a year later, in 1983, Ermoliev, Kaniovski and I published a collection of theorems in the Soviet journal *Kibernetika.*

Not too long after the appearance of this article, a Soviet colleague passed us a copy of an article by three US probability theorists, Bruce Hill, David Lane and Bill Sudderth, in the 1980 *Annals of Probability*, formulating and solving much the same problem. We were naturally disappointed. But as it turned out the Hill, Lane and Sudderth paper had solved a simpler, one-dimensional version of the problem; and we had solved the *N*-dimensional version. Moreover, our methods were different. We had used crude but powerful methods; they had used classic, but weaker methods. In the subsequent extensions of these theorems that we produced throughout the 1980s, we were able to borrow some of the Hill, Lane and Sudderth techniques and use them in combination with our stochastic approximation methods. Much of this work was carried out by mail. Some took place in Vienna. Four times I visited the Soviet Union, staying for periods up to a month. To keep up with this collaboration I was forced to learn a great deal of professional-level probability theory. Working with Ermoliev and Kaniovski was a pure source of joy.

The result of all this concern with probability was a general procedure for settling the selection problem in increasing returns problems. It would work by redefining each problem as a corresponding stochastic process, usually involving allocations or transitions among categories. The process itself might have multiple asymptotic states (or random limits), and one of these would be "selected" in each realization, not necessarily the same one each time. Thus the fundamental indeterminacy would become a probabilistic indeterminacy; and selection could be studied by examining the workings of the transient dynamics of the process. Often the non-linear Polya and stochastic approximation formats I had worked out with Ermoliev and Kaniovski could be applied; but sometimes other formats might be more appropriate. One paper here (Chapter 2) uses a random-walk formulation; another (Chapter 9) a master-equation format.

In casting around early on for examples of increasing returns at work within the economy I had become fascinated in 1980 with the economics of technology. The standard technology problem in economics was that of figuring out the economic circumstances under which a new, superior technology might replace an old inferior one, and how long this process might take. But from my engineering studies as an undergraduate, I was aware that a new technology normally came along in several different versions or design format. Thus if a new technology were replacing an old one these alternatives might well be thought of as in competition for adopters. Further, it seemed that learning effects would provide advantage to any version that got ahead in cumulative adoptions; and so the adoption process could lock in, by historical chance, to whichever version of the technology got a better start. It was clear that this "competing-technologies problem" was *par excellence* one of increasing returns and it seemed just right for the approach I was trying to develop.

In 1980 and 1981 I tried various formulations of the competing technologies problem. I gave a plenary address on it at the International Conference on Systems in Caracas, Venezuela, in July 1981, and it was received with enthusiasm. It took another year or two to reduce the competing technologies problem to a form I was satisfied with. I wrote it up finally as a IIASA working paper in summer of 1983.

In the meantime, in 1982 I had moved to Stanford. I was heavily involved in economic and mathematical demography, and spent much of the next three years reorganizing Stanford's efforts in demography. At Stanford, I met the economic historian, Paul David. He was extremely sympathetic to my ideas, and indeed had been thinking along the same lines himself for quite some time before meeting me. The introduction to his 1975 book *Technical Choice, Innovation, and Economic Growth* contains several pages on the connection between nonconvexity and historical path-dependence. Paul was intrigued at the prospect of an increasing-returns-path-dependence theory proper. Were there examples to go with this? I had been collecting papers on the history of the typewriter keyboard, and using the QWERTY keyboard as an example in papers and talks. Paul had thought of that, as had several others in the early 1980s. For argument he raised the standard objection that if there were a better alternative, people would be using it. I disagreed. We continued our discussions over the next

two years, and in late 1984 Paul began to research the history of typewriters. The result, his 1985 "Clio and the Economics of QWERTY" paper in the *American Economic Association Papers and Proceedings* became an instant classic. For me this paper had two repercussions. One was that the path-dependence rapidly became a familiar part of the thinking of economists; it was legitimized by a well-known figure and finally had a place in the field. The other, less fortunate, was that whatever Paul's disclaimers, for quite some time thereafter I was regarded by many as the man who formalized Paul David's ideas.

My 1983 technologies working paper (Chapter 2) received a great deal of attention, especially among economists interested in history and technology. But it did not do well at journals. In writing it up I had decided to keep the exposition as simple as I could so that the ideas would be accessible to the widest readership possible, even undergraduates. Many of my previous papers were highly technical, and several were in professional mathematical journals; and I saw no reason to dress the paper up in mathematical formalisms merely to impress the reader. I admired the lucidity and simplicity of George Akerlof's classic "The Market for 'Lemons'" and tried to write the paper at that level. This turned out to be a crucial mistake. The straightforward random-walk mathematics I used could not pass as an exercise in technique; yet the paper could not be categorized as a solution to any standard, accepted economic problem. The paper began an editorial and refereeing career that was to last six years. I submitted it in turn to the *American Economic Review*, the *Quarterly Journal of Economics*, the *American Economic Review* again (which had changed editors), and the *Economic Journal*. In 1989 after a second appeal it was finally published in the *Economic Journal.*

In early 1984 I began work on increasing returns and the industry location problem. I had been reading Jane Jacobs's *Cities and the Wealth of Nations* and had been greatly taken by her haunting accounts of places and regions that had got "passed by" historically in favor of other places and regions that had got ahead merely, it seemed, because they had got ahead. To prepare for working out a stochastic dynamics that would model industry clusters forming by historical chance under agglomeration economies I read a good deal of the German literature on spatial location. It appeared there were two points of view. Most authors, the better known ones mainly, favored an equilibrium approach in which industry located in a unique predetermined way. But others, usually obscure and untranslated, emphasized the role of chance in history and the evolutionary, path-dependent character of industry location over time. In the 1930s the path-dependence ideas, it seemed, had been largely abandoned by theorists. There had been no means by which to settle how one location pattern among the many possible might be "selected" and theory did not at the time accept indeterminacy. This problem was thus a "natural" for a probabilistic dynamic approach that could deal with the selection problem; and my resulting 1986 paper "Industry Location and the Importance of History" (Chapter 4) received attention at Stanford. But in the editorial process it met a fate similar to the competing technologies piece. After turn-downs from two mainstream journals, partially on lack of understanding that this was a legitimate problem for economic theory ("the paper would be better suited to a regional economics journal"), I finally managed to place it in *Mathematical Social Sciences* in 1990.

In looking back on these difficult times, I realize I gained much from them. I learned to persevere with my own ideas and not to look to current fashions for assurance. I gained friends in the profession who taught me about parts of economics I had previously not been aware of. I was forced, in self-defense, to learn new techniques. And I gained freedom. Not making much progress within the standard conventions of the profession, I became less concerned with them. This allowed a wider range, and encouraged me to learn from physicists and biologists.

Yet it is a sad commentary on our profession that my experience is not an isolated one. Other economists who have tried to pursue unusual directions have met with similar stonewalling. Certainly the profession needs to protect itself against crackpot ideas. And surely, we believe, if a new idea is worth anything it should be able to prove itself "fitter" than its competitors by surviving the selective pressures of journal refereeing. Yet perhaps the field has become too protectionist and inward looking.

The problem, I believe, is not that journal editors are hostile to new ideas. The lack of openness stems instead from a belief embedded deep within our profession that economics consists of rigorous deductions based on a fixed set of foundational assumptions about human behavior and economic institutions. If the assumptions that mirror reality are indeed etched in marble somewhere, and apply uniformly to all economic problems, and we know what they are, there is of course no need to explore the consequences of others. But this is not the case. The assumptions economists need to use vary with the context of the problem and can not be reduced to a standard set. Yet, at any time in the profession, a standard set seems to dominate. These are often originally adopted for analytical convenience but then become used and accepted by economists mainly because they are used and accepted by other economists. Deductions based on different assumptions then look strange and can easily be dismissed as "not economics." I am sure this state of affairs is unhealthy. It deters many economists, especially younger ones, from attempting approaches or problems that are different. It encourages use of the standard assumptions in applications where they are not appropriate. And it leaves us open to the charge that economics is rigorous deduction based upon faulty assumptions. At this stage of its development economics does not need orthodoxy and narrowness; it needs openness and courage.

My fortunes changed rapidly in 1987. The Guggenheim Foundation awarded me a Fellowship to study increasing returns in early 1987, and in April of that year, Kenneth Arrow invited me to come to a small institute in Santa Fe for a ten-day meeting in September that would take the form of a series of discussions between physicists and economists. I went; and from this much else flowed. The physicists there, particularly Phil Anderson, Richard Palmer and David Pines, immediately recognized the similarities between my outlook in economics and condensed-matter physics, and their endorsement did much to legitimize my work. The overall meeting succeeded enormously, and led to the idea of an Economics Research Program at Santa Fe. I was asked to be its first Director and accepted. The two years I spent at Santa Fe in 1988 and 1989 were the most exciting of my professional life.

At Santa Fe in September 1987 I had shared a house with John Holland, and had became intrigued—entranced—with his ideas on adaptation. These ideas seemed a long way from increasing returns; and indeed I did not particularly push research in increasing returns at Santa Fe in the first two years. Learning and adaptation, I believed were more important. But as I read into the literature I realized that where learning took place, beliefs could become self-reinforcing, whether at the Hebbian neural-synapse level, or in Holland classifier-system learning, or in learning in macro-economic problems. Thus I began to see a strong connection between learning problems and increasing returns, and as if to confirm this, much of the stochastic mathematics that applied to increasing returns turned out to apply also to learning problems. Although my recent work on learning is outside the scope of this volume, I have included a paper here that at least hints at the increasing returns connection with learning.

<p style="text-align:center">* * * *</p>

As of writing this, increasing returns are currently the subject of intense research in economics. Paul Romer's theories of growth are being connected with the international trade literature. Paul Krugman has taken up the industry-location-under-increasing-returns problem to great effect, and has done much to popularize it. Andrei Shleifer, Robert Vischny and Kevin Murphy's modern revival of the Rosenstein-Rodin "Big Push" argument has launched a renewed interest in increasing returns among development economists. Paul David and Douglass North have gone deeper into path-dependency and its meaning for economics in general, and economic history in particular. Timur Kuran is applying increasing returns to problems of social choice and political upheaval. Paul Milgrom and John Roberts have worked out a theory of complementarity. And Steven Durlauf and Kiminori Matsuyama are pursuing the stochastic, equilibrium-selection point of view. Several other first-rate economists are involved; and the subject, I am happy to say, is flourishing.

From time to time an economist will ask me where I am heading with my own viewpoint on economics. I used to believe I had no intended direction—that I was just following where the ideas led. But in reading through these essays I realize that from the first I have had a very definite direction and vision. The actual economic world is one of constant transformation and change. It is a messy, organic, complicated world. If I have had a constant purpose it is to show that transformation, change, and messiness are natural in the economy. These are not at odds with theory; they can be upheld by theory. The increasing-returns world in economics is a world where dynamics, not statics, are natural; a world of evolution rather than equilibrium; a world of probability and chance events. Above all, it is a world of process and pattern-change. It not an anomalous world, nor a miniscule one—a set of measure zero in the landscape of economics. It is a vast and exciting territory of its own. I hope the reader journies in this world with as much excitement and fascination as I have experienced.

W. Brian Arthur

Stanford, May 1993