ECONOMIC RESEARCH REPORTS

TOWARD A THEORY OF INDUSTRIAL POLICY-RETAINABLE INDUSTRIES

BY

Ralph E. Gomory
and
William J. Baumol

RR # 92-54 *December, 1992*

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

Toward a Theory of Industrial Pclicy - Retainable Industries

Ralph E. Gomory[1] and William J. Baumol[2]

Introduction

Ricardian international trade analysis, even in its modern forms, focuses upon industries whose technology offers no economies of large scale production. The assumption of scale diseconomies seems appropriate for a Ricardian world dominated by agricultural and textile products, but there are significant products and industries in the modern world that are better described by a generalized notion related to economies of scale which, for reasons that will become clear, we will call "*retainability*". As we will show, the Ricardian world and the world of generalized scale economies yield profoundly different conclusions:

1. In a Ricardian economy we expect the international trade equilibrium to be unique (though there can, of course, be exceptions). A world economy dominated by retainable industries is characterized by a vast number of stable equilibria, even for a two country model on the order of at least 2^{n-1} where n is the number of goods. With even a moderate n, the number of equilibria is so large that a method of aggregating them is necessary for the analysis. In this paper therefore we discuss aggregates of equilibria which we call *regions of equilibria.*

2. In the Ricardian model the loss of an industry by nation j is normally compensated by the gain of other industries, in accord with shifting comparative advantage. Retainability, in contrast, implies that the loss of an industry by Country j need not be accompanied by the

---

[1] Alfred P. Sloan Foundation.

[2] Professor of Economics, New York University, and Director, C. V. Starr Center for Applied Economics, N.Y.U.

gain of any others. Thus, the loss of an industry by a country will normally decrease its relative national income, while the gain of an industry will increase it. This opens up the possibility of a country enriching itself by the acquisition of industries while its trading partner is steadily impoverished by their loss.

3. Ricardo's world is characterized by a unique Pareto optimal equilibrium that also provides mutual gains from trade. Under retainability, many equilibria violate comparative advantage and are economically inefficient. In fact whole regions of equilibria are vastly inferior to others for all of the countries involved, and many are even worse than autarky for one of the countries.

4. Under the Ricardian analysis the market can, consequently, be trusted to yield desirable and efficient results. Where scale economies or retainability are widespread market forces still produce an equilibrium, but that equilibrium is only one of a vast array of possible equilibria forming a region of equilibria. It is the location of a particular equilibrium point in the region that determines its desirability for one country or the other. That location can be the product of historical accident or of more deliberate national policies.

The pertinence of these results, both to theory and to policy, should be clear.[3]

I. The Concept of Retainable Industries

We use the term "retainability" to lump together a number of loosely related phenomena, all pertinent to our analysis. We will, of course, include the traditional economies of scale and the attendant cost advantage which a larger scale industry, with a great variety of specialized support services, will, in some industries, have over an established but smaller producer. But we will also include the other advantages which a country that is actually engaged in production has over countries that are non-producers or very small scale producers. These advantages can include knowledge gained by experience -- actually knowing how to produce a complex or technologically advanced product, access to a distribution channel that will only take a few

---

[3] These results are complementary to those of an already very fruitful literature on the role of scale economies in international trade. See the seminal contributions of Mathews [1949/50], Meade [1952], Chipman [1965], Kemp [1969], Ethier [1979,1982], Helpman and Krugman [1985], Krugman [1991], and Grossman and Helpman [1991].

distinguishable products, etc. A current example of this is the Japanese advantage in liquid crystal displays. These displays have taken a major technological leap forward with the introduction of thin-film-transistor liquid crystal displays. Items using this new technology are made only in Japan today, and the specialized knowledge required to produce them defect free on a large scale makes it unlikely for a long time that others will be able to enter this field successfully, for they will have to make a large sunk investment in trial and error learning to reach a state of any production at all, while the Japanese industry has already incurred that sunk investment, has emerged on the other side and is still learning.[4] This advantage of the "first mover" is a type of benefit of relative scale common today in a world of rapid technological progress, that also distinguishes our world from the Ricardian one.

In our language, *a "retainable" industry is one that can be held onto by a country, despite lower wages in other countries*, because of any one, or a combination of, these factors. One of the central implications of our analysis is that a nation can increase its relative income by increasing its share of the world's retainable industries.

Retainable industries have significant implications for a world of relatively free trade in which high wage and low wage countries are in active competition with one another. It will be greatly to the advantage of a high wage country to specialize in retainable industries. The role of such industries is also of immediate relevance to the situation of the industrialized countries.

Much of our discussion will be expressed in terms of economies of scale, traditionally defined. This route will be followed because it simplifies exposition. However this case restricts unduly the applicability of the analysis, so that at the end we will reformulate the implications in terms of the more generally applicable concept of retainability. In this sense Gomory [1992] and Gomory and Baumol [1992] although couched in terms of economies of scale, can also be thought of as the beginnings of a theory of retainable industries.

II. Summary of the Analysis.

Before turning to a more general analysis we will give an example, which though exceedingly simple, illustrates a key point. Let us consider a miniature economy composed of

---

[4] Alternatively, a Japanese partner may be required for successful entry into the field.

two countries, Japan and the U.S., producing only two commodities, semiconductors and airframes. We will assume retainability in both. As we will see, as the analysis develops, it is extremely helpful to focus upon *perfectly specialized equilibria*, that is, equilibria in which no commodity is produced simultaneously in two or more countries. One such equilibrium outcome is that in which Japan produces all the semiconductors, and the U.S. produces all the airframes. This is a stable outcome because it is hard for the U.S. to enter the semiconductor field, given the assumed economies, and the same applies to hypothetical Japanese entry into airframes.[5] *However another and equally legitimate and sustainable equilibrium is one in which Japan produces all the airframes and the U.S. all the semiconductors,* and for the same reason, the non-producer can only expect a negative profit from entry into the industry on a small scale against the entrenched producer. This is an example of retainability. Either equilibrium will be sustained by normal market forces. Which of them is actually realized, in the absence of some non-market intervention, is, in the model, a fortuitous consequence of the selected starting point, or in real life, of history (on this, cf. Krugman [1991]). Moreover, it should be obvious that one of the two specialized equilibria in this case is virtually certain to violate the pertinent comparative advantage conditions, say, if Japan is the relatively poor airframe producer and the U.S. is, consequently, the relatively inferior semi-conductor producer.

If we deal with many goods instead of just semiconductors and airframes, the number of such perfectly specialized equilibria becomes very large, since there are many possible ways in which subsets of the set of industries can be assigned to each of the two countries. With two goods we had two specialized equilibria (one choice of producing country for each good). With n goods we will have $2^{(n-1)}$ because for each good there are two choices of producing country.[6] As will be show presently, despite the large number of variables in the model, it is possible to represent the equilibrium points in a two-dimensional graph that will be described next. Two

---

[5] We will see later that all the conditions for a competitive equilibrium, including zero economic profit for producers, negative profit for non-producers, and expenditure on each good equal to price times quantity produced, can in fact be met by this outcome. [See Section III].

[6] For simplicity we are assuming that each country is capable of producing some quantity of any of the goods. Also while we discuss here specialized equilibria, we will examine the role of non-specialized equilibria later in this paper.

remarkable features of this graph are (a) the fact that the equilibrium points do not fall haphazardly throughout the diagram but, rather, that the *equilibria all lie in a well-defined region* which (with a sufficient number of commodities) they tend to fill up; and (b) the boundaries of these *equilibrium point regions have a characteristic shape of considerable economic significance*. This will enable us to deal with the enormous number of equilibria as an aggregate.

In this graph, any equilibrium is described by the values of three of the pertinent variables (Figure 1). The first of these variables is obtained from our models by calculating for each equilibrium point a figure for the level of national income, Y_j, in each Country j. From this we can construct an index of the *relative* income of Country 1, which we will call $Z_1 = Y_1/(Y_1+Y_2)$. We plot this variable on the horizontal axis of the graph, so that axis extends from $Z_1 = 0$ to $Z_1 = 1$, that is, we go from the case where Country 1 obtains none of the two nations' total income, at one extreme, to the other polar case in which all of that income goes to Country 1 alone.

Our models will also show how much of each good will be consumed in each country in a particular equilibrium, so that if we assume the existence of a utility function for each country, j, we can use that function to calculate U_j, the utility that the equilibrium in question yields to that country.[7] Thus, each equilibrium is characterized in our graph by the values of three variables: Z_1, U_1 and U_2. Since the utilities of the two countries cannot be compared directly (they are *ordinal* utilities) we need two different axes along which to measure U_1 and U_2. We use the left-hand vertical axis to measure U_2 and the right hand axis to measure U_1. Thus, each equilibrium is represented by two points, P_1 and P_2, in our graph. The one, P_1, reports (Z_1, U_1), the utility it gives to Country 1, as a function of its share of world income. The other point, P_2, represents (Z_1, U_2). Point P_1 for a given equilibrium must lie either directly

---

[7] It is not necessary for our discussion to interpret U_j as an aggregative utility function for Country j, a concept that has, understandably, elicited some controversy. Rather, one can legitimately think of U_j as a weighted average of the physical quantities of goods consumed in that country. For example, using q_{ij} to represent the quantity of good i consumed in Country j, then such an index is provided by the Cobb-Douglas expression $U_i = \pi_i q_{ij}^{\alpha ij}$, where $\Sigma\, \alpha_{ij} = 1$.

below or directly above P_2 because the value of Z_1 is given once the equilibrium under consideration is selected. One final feature of the graph needs to be noted. On the U_1 axis we have plotted a point labelled AUT_1. This represents the utility that Country 1 would obtain in a world of autarky, that is, one with no trade between the two countries. Point AUT_2 on the other vertical axis has a perfectly analogous interpretation.

Several general properties are illustrated by Figure 1. As was already noted, one of its remarkable features is that the $2^{(n-1)}$ equilibria, any of which will be sustained by market forces, do not appear randomly throughout the diagram, but fall *in and tend to fill up a well defined region*, bounded by the four well defined curves, labelled $U_{1\,max}$, $U_{1\,min}$, $U_{2\,min}$ and $U_{2\,max}$ in the graph.

If these characteristics hold for many models, and we will show that they do, it becomes necessary to rethink what it means to leave things to pure market forces. Market forces can move the world economy to *any* one of the equilibria, which is to say, almost anywhere in a large region, yet the location of the outcome matters enormously to the countries involved under scale economies. The market can still be expected to drive the economy toward some equilibrium, but here the equilibrium is far from unique, so it is the identity of the *equilibrium* that is actually approached, or equivalently, *of the part of the region of equilibrium points that the country is in*, that is really decisive. But there is nothing to lead market forces consistently toward any particular part of the equilibrium region.

III. The Nature of The Equilibria in the Model

Let us now briefly outline the formal model and report some of its additional properties.[8] Then, the section that follows will offer intuitive explanations of the properties already reported. The two-country model is composed of a utility function for each of the countries, a demand function (derived from the utility function) for each good by each of the countries, a production function for each good for each country (with each good assumed to be produced independently

[8] For fuller discussion of the materials in this and the following sections, and rigorous derivation of results, see Gomory [1992] and Gomory and Baumol [1992].

of any other) and a fixed supply of input (labor) for each country.[9]

The equilibrium conditions consist of four requirements: (1) that prices be such that the demand for each commodity by the two countries together be equal to the quantity of that good supplied; (2) that no good's production in either country be able to yield a positive profit, and that if good i is actually produced in Country j its output must be such that this process yields zero profit[10]; (3) that relative wages be such that the quantity of labor used in total in each country is equal to the available quantity; and (4) that payments of the two countries must in balance.

These four sets of equilibrium requirements yield the quantity of each good consumed in each country, the quantity of input used in the production of each good by each country, the price of each good, the relative wage rate, the exchange rate, and the relative incomes of the two countries.

As described so far, there is nothing very unusual about the construct. However, we will now focus upon specialized solutions and we will see that matters depart a bit from their usual pattern. Before proceeding with the analysis we will make a few remarks about specialized and non-specialized solutions in the presence of economies of scale.

Let us assume that several countries are initially successful producers of some product, X, that is characterized by scale economies. Then, the costs in those countries must at that point be similar. In these circumstances, if any one of the countries, Country A, fortuitously or by

---

[9] Specifically, the utility functions are assumed to be of Cobb Douglas form, and the production functions $f_{i,j}(l)=y_{i,j}$ have economies of scale and zero derivatives at the origin. y_{ij} is the output of good i in Country j, l_{ij} is the labor devoted to its production. We also assume that each country, in autarky, is capable of producing some positive quantities of each of these goods. Particular choices of the production function can be significant in facilitating the computational process. But the logic of the relationships indicates that our qualitative conclusions are robust, and do not depend on the functional forms utilized in the calculation of the equilibrium values of the variables. This is equally true of the premise that despite scale economies, profits are zero throughout the economy.

[10] This is not the place to review the 20 year old debate on the compatibility (first assumed by Marshall) of scale economics and ubiquitous zero profits. It is enough to note here that the generalized concept of retainability is consistent with the presence of a number of horizontal competitors whose rivalry drives economic profits to zero.

design, manages to expand its share of world output of X its costs will clearly fall below those of the other nations. It will then capture even more of the world market by virtue of those low costs, and the larger the proportion of the market A acquires the greater its cost advantage will grow. The process can only cease when that country has become the exclusive producer of X. Thus, unspecialized equilibria will tend to be unstable, and for the obverse reasons, specialized equilibria will be locally stable. That is, if any other country, B, seeks to break into the market for product X, it cannot hope to succeed by entering the field with an initially small output of X. Scale economies mean that B can only hope to achieve a change from the equilibrium in which it is totally excluded from the X market by entering that market with a very large output of X on the day its X industry opens for business, and it is this property that connects economies of scale with retainability. Thus, in a scale economies world, while large moves in a country's output of a good from whose production it was previously excluded can change the equilibrium, small moves can be expected to fail to draw the situation permanently away from the initial equilibrium. These remarks suggest the following two propositions:

Proposition 1. Each and every perfectly specialized solution of the international trade model with scale economies, two countries and n commodities is also an equilibrium solution. A simple dynamic analysis of Marshallian type can then also be used to show the companion result.

Proposition 2. Each and every perfectly specialized equilibrium is locally stable.

A careful analysis of the equilibrium equations will in fact lead rigorously to the proof of these two propositions provided we formulate more carefully some of the properties of economies of scale and of the abilities of the countries to be actual producers of each and every good.[11]

When we assume, as we do here, that our production functions are such that for sufficiently small quantities of production the unit costs are prohibitively high we should remember that this can result either from economies of scale in production, or, equally plausibly, from the years that must be spent on learning and training before anything at all emerges.

---

[11] Adequate assumptions are that the production functions, f, have zero derivative at the origin and that in autarky each country can produce positive quantities of every good, however small. But the latter premise is not necessary for our results.

Our two propositions now give us, even in relatively small models, a huge number ($2^{(n-1)}$) of different specialized equilibria mentioned earlier. These equilibria can be shown to possess two further important properties.

Proposition 3. All these specialized equilibria lie within a well defined region, such as the one in Figure 1, whose upper and lower boundaries can be found by a simple linear programming type of calculation, without obtaining the equilibria themselves.

Proposition 4. The entire region between the upper and lower boundaries becomes filled with equilibria as the number of industries increases. More precisely, if p is any point in the region, then for n sufficiently large, there will be an equilibrium point arbitrarily close to p.

Our discussion so far has emphasized specialized solutions and has emphasized that they are likely outcomes. Nevertheless the theory does in fact also shed considerable light on non-specialized solutions.

Proposition 5. Every non-specialized equilibrium, whether stable or not, lies under the upper boundary of the region defined by the specialized equilibria.

Thus, our emphasis on specialized equilibria has two justifications. One is that economies of scale, left entirely to themselves, will tend to produce them. The second is that, non-specialized equilibria, whether stable in the model or not, are contained below the upper boundary of the region that we analyze *and whose size and shape are already determined by the specialized equilibria.*

IV. Intuitive Explanation of the Model's Policy-Related Properties; Gaining and Losing Retainable Industries.[12]

This section describes the policy implications of the retainable industries case. We will see, with the aid of our graph, that these circumstances reduce the reliability of the market mechanism as a generator of socially-optimal equilibria, that this case provides a wider range

---

[12] Throughout this paper statements about the gain or loss of industries are to be understood in the sense of comparative statics. That is, they entail a comparison of two alternative equilibria, in one of which Country j holds more retainable industries than in the other, all other things remaining equal. Thus we do not refer to a "before" and "after" comparison--an intertemporal move entailing the acquisition of industries that Country j did not have before.

of possibilities for clashes in the economic interests of nations and, consequently, that it opens up new ways in which one country can benefit at the expense of another.

For reasons that will become clear presently, the region of Propositions 3 and 4 generally has the shape displayed in Figure 1. For Country 1 it rises from zero when $Z_1=0$, i.e., when Country 1's relative national income is zero, increases rather steadily and then, after a certain point, turns down again.[13]

Policy Implications of the Characteristic Regional Shape.

Let us discuss the policy implications of such a shape for the equilibrium region. Consider the utility and relative national income of Country 1 at a specialized equilibrium point E_a at which it is the producer of a set of goods S_a and compare that with its utility and relative national income at another specialized equilibrium E_b at which it is still the producer of the goods in S_a but in addition is the sole producer of one additional good. We would expect that country's income to increase since it now has as part of its national income the world expenditure on the new good. Similarly the relative national income of Country 2 should decrease as it has lost that portion of income.[14] Thus, since in E_b Country 1 holds a larger number of industries than at E_a, we expect E_b to lie to the right of E_a. That is, at E_b the relative income of Country 1, Z_1, will be greater than at E_a.

We will call such a shift, from E_a to E_b "capturing a retainable industry" or, from the point of view of Country 2 "losing a retainable industry". Clearly, a succession of such changes leads to steady increases in the relative national income of Country 1 and decreases in the relative national income of Country 2. This, in turn, will generally lead to increases in *utility* for Country 1 and decrease in *utility* for Country 2 up to the peak in the upper boundary, after

---

[13] In Gomory and Baumol [1992] we show that this regional shape does necessarily hold for a wide class of models.

[14] With a given labor force and full employment, the outputs of Country 1's other products must fall and so their prices must rise. The argument in the text assumes that demands are not so elastic as to cause a decline in revenue in the old industries greater than the gain in revenue from the newly acquired industry.

which it is likely to result in decreases for *both* countries (Figure 1). We say generally, because there is the possibility that E_1 can be near the upper boundary and E_2 near the lower boundary, or the reverse, but on average that will not be the case.

In the central subregion (Figure 1) between the peak points A and C the interests of the two counties are opposed; an increase in utility for Country 1 is generally accompanied by a decrease in utility for Country 2. However, to the left of A or to the right of C, their interests generally coincide.

The same argument applies to any shift from one equilibrium point to another that increases relative national income, whether or not the points are specialized equilibria.

Explanation of the Characteristic Regional Shape.

All of this follows from the characteristic shape of the equilibrium region. We will now try to show that this characteristic shape is, in turn, the consequence of some simple economic relationships.

Consider the circumstances of Country 1 near $Z_1=0$. The reasons for the positive slope of Country 1's upper utility boundary in this region seem obvious, since its relative income rises as we move to the right. Still, it is worth spelling out the details of this part of the story. Near $Z_1=0$ Country 1 is the producer of very few goods that are traded internationally, and the importer of many. Since it makes few goods, its national income, which is the sum of world expenditures on those goods, will be low and its wage, which is that expenditure spread over the labor force, will be low. Because of its low wage it is potentially the cheaper producer of many more goods than it is actually making, if only it could get over the difficulties of entry posed by economies of scale (or of learning, or of sunk costs). If by some means Country 1 can capture an additional industry there are several effects, mostly benign from the point of view of Country 1 and detrimental from the point of view of Country 2. First, the relative wage in the high wage country, Country 2, will have decreased, or, equivalently, the exchange rate will be lowered, since Country 1 is producing more and importing less. This increases Country 1's share of each good produced as Country 1 is now a larger slice of the total world demand for each good. Second, the cost of the goods from the newly acquired industry are likely to be reduced, because of the low wage in Country 1. Third, all the goods imported into Country 1 from Country 2 are likely to become cheaper, since Country 2 will have its labor distributed

over fewer industries than before[15], and therefore will devote more labor to each industry --- increasing outputs and cutting costs and prices. All of these influences tend to increase the utility of Country 1.

The negatively sloping portion of Country 1's upper utility boundary to the right of point B in Figure 1 is more unexpected. We will explain it in two different ways. First, note that as we approach the extreme right, and Z_1 approaches unity, Country 1 has become the producer of almost everything, Country 2 makes little, has a low national income, and matters little to Country 1 either as a producer or as a market. Country 1 is substantially alone in the world as it would be in autarky. For this reason the utility value approached on the right in Figure 1 is the Country 1 autarky level of utility, a decrease in utility from point B for Country 1 as that country gives up even more of the possible gains from trade.

Alternatively, the negative slope of the right-hand end of Country 1's upper boundary can be explained as follows: As we move to the right the relative wage in Country 1 must go up, and its labor is now spread over more industries so that the goods it makes will have become more expensive, even in Country 1. Near $Z_1=0$, Country 1 makes few goods, while Country 2 makes many, so the beneficial effects of Country 1's gain outweigh the detrimental consequences, for *both* countries. But if Country 1 acquires industry after industry, its relative wage rises, it changes from the cheaper to the more expensive producer of the newly acquired good, its imported goods continue growing cheaper but there are few of them, while the goods made in Country 1 itself, which are now most of the goods it consumes, become more expensive. Clearly, then, at some point the utility to Country 1 of capturing more industries turns negative, and this is the behavior that we see in the figures.

This analysis explains the main features of the model and can be made more explicit in a detailed mathematical description of the model as is done in Gomory and Baumol [1992]. However, there is one further aspect of the shape that is worth noting. We see that there are many equilibria in these figures that are above the autarky level, just determined, but also many

---

[15] Note that at this point we depart very strongly from the diseconomies model. In that case the lower wage could result in the recapture of some industries by Country 2 with its new lower wage. Recapture here is ruled out by retainability.

that are below it. Thus many of the solutions can entail losses from trade for at least one country.

Implications About the Market Mechanism.

We should emphasize that our central construct, that of a *region* of equilibria, has very strong consequences. Because every point in the region is, in a large model, near an equilibrium point, it is a reasonable approximation to say that *every* point represents an outcome that can be produced by market forces.

Now one of the prime virtues of the market mechanism in a world of scale diseconomies is its tendency to approximate optimality, if there are no substantial externalities and if its workings are not interfered with either by government intervention or deliberate distortion by private interests (for example, through attempts at monopolization). In the absence of externalities, if diminishing returns are everywhere present and entry and exit are completely free, there will generally be a unique equilibrium. That equilibrium will be Pareto optimal, and hence efficient, and market forces can be relied upon to drive the world economy toward that equilibrium arrangement.[16]

In contrast, where retainable industries play a widespread and substantial role, as we have seen, there will normally be a large multiplicity of possible equilibria, all of them stable, and many of them highly detrimental to some or even all of the participating economies. One cannot predict in any general way toward which one of these equilibria the market will direct the world economy. That depends on the happenstance of history, the political and economic predispositions of governments and other fortuitous influences. Thus, there is no more reason to expect the market forces to move the world economy toward an equilibrium point that performs well in promoting the global public interest, or that of any particular country, than to anticipate that those forces will land the world in an equilibrium point that is highly detrimental

---

[16] In Baumol and Gomory [1992] we discuss the efficiency of the various equilibria, we introduce the notion of "local efficiency" and conclude that while the equilibria in our model are rarely efficient, they are always locally efficient. Thus market forces continue to have their traditional virtue, the ability to produce a stable outcome that is optimal locally. However, in the presence of economies of scale there are many such locally optimal points and market forces do not choose systematically among them and so do not necessarily lead to one that even approximates global efficiency.

to the public. Thus, the absence of interference with the workings of the market mechanism, whether by private groups or by government agencies, can no longer be considered to be invariably beneficial. In short, in a world pervaded by scale economies or by other retainable industries, non interference with the solutions provided by the market can not legitimately be assumed to be an unqualified virtue.

V. Some Scale Diseconomies and the Limits of Retainability.

Of course, the economy of reality is not composed exclusively of industries that are retainable or have scale economies, and no country ever ceases producing altogether, as does one of the countries in our model, as described so far, when we reach either of the vertical axes. To bring our construct more closely in line with reality it must be adapted to include some industries with scale diseconomies. Thus, even if one of the countries should succeed in capturing all of the scale economies industries, the other country will still be able to produce and trade in the commodities whose production does not entail scale economies. In addition, any economy is always in a position to turn out the many non-exported goods and services which, in reality, constitute a very substantial proportion of the GDP of any country.

Fortunately, it has been possible to extend the model to encompass the case of scale diseconomies. This modification yields two main results, neither of which should be considered particularly surprising.

First, toward, say, the righthand end of the Z_1 axis where Country 2 has been deprived of every scale economies industry, Country 1 may nevertheless continue to produce some of the scale diseconomies goods or services. Thus, even in this case, Country 2 will generally not be left alone in any diminishing return enterprise, to serve as sole producer immune from foreign competition. This case is plausibly interpreted to characterize the relationship between the industrialized countries and the LDCs. The latter produce few, if any, of the world's scale economies products. Their focus on products of agriculture, forestry and mining surely indicates that. But even in these fields they encounter competition from the industrialized countries, all of which have their agricultural sectors, and many of which provide lumber and minerals as well.

There is a second major observation yielded by the inclusion of products with scale diseconomies. As the share of diminishing returns industries among traded commodities grows, the range of Z_1 values over which the region of equilibrium values extends grows increasingly smaller. That is, as the share of diminishing returns industries grows from zero to 10 percent, then 25 percent, then 50 percent, (etc.) of the total, the region of specialized equilibrium points, which initially extends from one vertical axis to the other, narrows increasingly (Figure 2 shows a case calculated explicitly from one model). In the limit, as scale diseconomies goods and services approach 100 percent of the world's traded commodities, the shrinking region of equilibria approaches a single point, the unique full-employment equilibrium point that neoclassical analysis has taught us to expect.[17] In a regime of universal perfect competition without externalities this equilibrium will also have all the desirable welfare properties that Arrow-Debreu theory has led us to expect.

A second aspect of reality that can also be taken into account is limits on the retainability of an industry. The model can easily be modified to allow, in several different ways for an ability to jump over the impediments to entry if the driving force of wage differential is great enough. The size of impediment that can be overcome can become a parameter of the model, or one can even have industry specific parameters. The parameter values will, among other things, reflect the effect of government policy toward industrial investment or toward investment

---

[17] We do not need to assume that any industry is characterized by scale economies for every possible output level. Rather, economists tend to think of the total cost function of, at least, the firm as characteristically manifesting scale economies up to a point, thereafter exhibiting linear homogeneity for some greater or lesser interval ("flat bottomed average cost curves") which, finally, is followed by a region of scale diseconomies. Such a scenario can entail a cost function for the industry with an initial range of scale economies and then constant returns to scale. For, then, further increases in market demand, rather than by being met by expansion of the incumbent firms into their region of diminishing returns, can be satisfied by the entry of new firms, most of them producing at similar costs, as they operate at or slightly beyond the output levels that are required for minimum efficient scale.

For the purposes of our analysis, however, it does not matter that if firms or industries grow sufficiently large, they eventually leave the region of scale economies, so long as the borderpoint at which scale economies are exhausted requires a sufficiently large volume of output. For the essence of our analysis rests on the cost advantage of having a relatively large position in some industry, and the consequent difficulty of entry into it on an (initially) small scale.

in particular industries. The effect of getting over these impediments is similar to the effect of having diseconomies industries. There is a steady contraction of the regions toward the classical solution as the barriers to entry become easier and easier to overcome.

In Figure 3, also calculated from a particular model, we can see this effect. The parameters that we use here (v_1 for Country and v_2 for Country 2) reflect the ability to enter in the following way: Country 1 can take over an industry held by Country 2 if, given a fraction v_1 of the world market, and its low wage rate, Country 1 can produce as cheaply as Country 2 when the latter serves the entire world market. In Figure 3a $v_1 = .2$, i.e 20%, and $v_2 = .05$ or 5%. Other values of the parameters are illustrated in the other parts of Figure 3. We see not only the contraction of the regions as the v_j increase but also the effect of the difference in the ability to enter, in Figure 3a, Country 2, whose industries are less retainable, does relatively poorly, Country 1, whose industries are more retainable, does far better.

In this part of the analysis a good deal depends on the shape of the production functions, and the sensitivity to parameter values varies greatly with that shape. However the general notion of contraction to the classical equilibrium is always there, as well as the advantage, say to Country 1, of being able to enter Country 2's industries more easily than the Country 2 can enter Country 1's.

VI. Retainability vs. Scale Economies.

It is useful to divide retainable industries into those characterized by scale economies in the standard sense of the term (call them "E-Retainable industries") and those that are not ("N-Retainable industries"). Both of them raise the same general policy issues. However, in a world populated with scale economies industries it is clear that market forces are apt to drive the world economy toward some perfectly specialized equilibrium, since the closer the economy is to such a point the greater the cost advantage that is enjoyed by the largest producer, and so the stronger the forces making for expansion of the large producer and contraction by the small. Put the other way, in a world of E-Retainable industries unspecialized solutions are likely to be unstable since any gain in output by one of the producers can be expected to inaugurate the mechanism pushing toward specialization.

In a world of N-Retainability, in contrast, nonspecialized equilibria may well be stable. N-Retainable industries may well be characterized by average cost curves that, after an initial range of rapid productivity growth, tend to flatten out. What matters for retainability is not the overall shape of these curves but rather the extent that they provide an impediment to entry. In an industry whose technology is evolving rapidly, a nation whose firms in that line of activity or in closely related lines that succeed in getting a substantial head start on everyone else, often acquire skills, experience, equipment and ways of thinking that render that industry quite defensible against assaults by foreign entrants.

For N-Retainable industries every specialized arrangement still constitutes a stable equilibrium. For the term refers to a state of affairs in which, when one country monopolizes some industry, i, other countries will find it exceedingly difficult to enter the market, and if they try to enter it on a small scale, they are apt to be at an extreme disadvantage.

Thus, N-Retainable industries yield the same immense set of perfectly specialized equilibria that E-Retainable industries do, they entail the same types of conflict in the interests of trading nations, the same possibilities for capture of industries, and the same prospect that the world economy will happen to end up at an equilibrium point where there are losses from trade for at least one of the countries.

N-Retainable industries can be modeled by using a production function that has a strong initial economy of scale followed by small or no economies of scale for larger values of the input. While non-specialized solutions in such a model are usually, but not always unstable, the forces producing instability are so weak that additional forces present in the real world, but not in such a simple model, can easily stabilize them at a point where both countries are large scale producers. Examples would be different transport costs to different markets, slight product differences that appeal to one class of consumer slightly more than to another etc. Similarly, in a many-country model, there may be several producing nations for any commodity, i. Though they are at a nonspecialized equilibrium among themselves, there is still a strong impediment to entry by those that are not producers. This is, of course, what we generally see in reality.

Econometric studies suggest that there are in fact many industries with scale economies. But the documented scale economies may only be local. We have no way of knowing whether

those scale economies would (or would not) be exhausted long before one nation's output of some such good i expanded sufficiently to approach total world output. In contrast, reports of persons with experience in industry support the casual observation that N-Retainable industries are fairly common.

VII. Toward Policy.

The main policy implications of our analysis have already been mentioned. First, where industries of the types described are present, it does become possible for a country to gain or lose the products of those industries for itself without, in implicit exchange, giving up other products to the nations from which the industries were captured. Second, government can, at least in principle, play a useful role in the struggle for such industries. Third, the industries that may be the most promising targets of such "industrial policy" are not characterized primarily by their prospects for growth (though promise of growth is, of course, desirable) but by their retainability, which may be the result of scale economies and the need for a heavy sunk investment, or it may have other, more technology or knowledge oriented, causes. Indeed, the instinctive preference that is often expressed for "high tech" industries may come from an instinctive appreciation that they may provide both retainability, either through knowledge, technology or scale economies, and at the same time provide growth.

Finally, the analysis implies that the "capture" of additional industries is not always desirable for the country that does so, and that consequently, the long run interests of both the industrialized and the less developed countries may be served by nurturing some scale economies industries in the latter.

At this stage of our discussion none of these three conclusions should require any lengthy explanation. Obviously, the policy directions they suggest are far from straightforward, and our theoretical analysis by itself is insufficient to indicate the details of the appropriate programs. We will, consequently, make no attempt here to explore any such policy measures in any detail, simply because we are in no position to do so. However, we will permit ourselves occasional remarks that go beyond what the theory supports. Let us, then, briefly consider our four conclusions in turn.

1. **The Gain or Loss of Retainable Industries.** Our analysis has shown that the different specialized equilibria are characterized by differences in the products they assign to the trading countries. One equilibrium, call it A, can assign a large share of the world's increasing returns industries to Country 1 and few to Country 2, while the reverse can be true of another equilibrium, B. Thus, the change from equilibrium A to equilibrium B clearly entails the shift of a number of industries to Country 2 from Country 1. Since B, like all the specialized equilibria, is stable, the capture need not be a very transitory affair. It is also clear that Country 1 will end up a net loser of industries in this transition process. There is no automatic replacement of its departed industries by others, as there would be in a world of scale diseconomies, as a result of market forces working in accord with comparative advantage. In short, where scale economies and sunk costs are present the capture of industry is very possible, as the recent history of the Far Eastern "Tiger Economies" suggests. Moreover, the positively sloping segments of the upper and lower utility frontiers in the graphs show that, at least over some range, it can be beneficial to the country that does so, and harmful for the country that loses it.

2. **Efficient Instruments for Industrial Policy.** Advocates of industrial policy have proposed for the purpose a variety of tools, many of which have actually been used by governments that have pursued such programs. Subsidies, tariffs, quotas and special pricing rules are among the means considered. Economic analysis suggests, however, that some of these means may constitute serious impediments to economic efficiency which inhibit production and impose unnecessary costs upon the public. To see which instruments promise to deal with the matter efficiently we must consider *why* the market is apt to fail to produce a global maximum in a world with many retainable industries.

At first glance, a complete explanation would seem to rest on the distinction between global and local maxima. Market forces are essentially local in nature, the firm reacts to prices as they are, to the cost of labor as it is, etc. and economic equilibrium is defined as a balance among these forces, so it is fundamentally local in nature. There is nothing in the definition of equilibrium that either asks or answers the question, can there be a better equilibrium somewhere else? Although this is true in a very general sense, it does not exclude the possibility that, for our particular models, a more explicit explanation can be given.

We turn naturally, as an explanation of market failure in the retainability case, to the difficulty of obtaining large amounts of capital for a risky, and often sunk, investment with the private risk possibly greater than the social risk. Equivalently, for the N-Retainability case, there is a long term commitment to learning or training or development of infrastructure, again for an uncertain return.

But beyond these considerations, which are real and widely recognized by economists in many other contexts, is something which is peculiar to this model, the role of *pecuniary,* rather than technological, externalities.[18] The market may fail to do the job because a large portion of the benefit generated in a country by its acquisition of a new industry takes the form of an externality for which investors in the industry are not fully compensated. The gain of an industry by Country j benefits its general populace by moving the exchange rate in its favor, raising the relative wages of its workers and thereby increasing its real income. So far as investors in this industry in Country j are concerned such benefits are genuine spillovers from which they themselves do not benefit. Consequently, the heavy investment necessary to acquire a retainable industry may characteristically be unprofitable for private investors, even though it is thoroughly justified in terms of the well being of Country j. Thus, in Adam Smith's words, the investments required to compete in such an industry "...though they may be in the highest degree advantageous to a great society, are, however, of such a nature, that the profit could never repay the expense to any individual or small group of individuals..." (*Wealth of Nations* Book V, Ch. I, Part III). It is true that the external benefit to the inhabitants of Country j comes at the expense of those of the other country that stands to lose the industry in question. Nevertheless, it is clear that the free market is likely not to elicit all of the private investment that is called for from the point of view of the narrow self interest of the prospective gainer country.

---

[18] In the standard model, of course, pecuniary externalities are not a source of market failure and are not a source of welfare loss for society as a whole. But, because pecuniary externalities entail changes in prices, they can (and presumably normally do) lead to some redistribution of income or wealth, entailing a welfare loss for one group and a welfare gain for another. That is the case here. Failure by Country 1 capitalists to invest sufficiently in industry i to enable them to compete successfully in that line of activity is detrimental to the workers in Country 1 and beneficial to those in Country 2.

Such causes of market failure have, of course, been explored extensively by economists in other contexts, and the literature suggests that the most efficient means to improve matters in such circumstances is for the public sector to intervene and to adjust the price of the pertinent commodity, the cost of capital, or to help with the acquisition of knowledge or technology and by helping to build the industry infrastructure. Where investment is inadequate it is, arguably, the price of capital that is out of line with the social interest and, in any event, adjustment in this price need entail no direct controls, trade barriers or other unnecessary impediments to the working of the price mechanism.

Government can influence the price of capital directly by acting to reduce interest rates or by providing insurance against risk. These are, reportedly, instruments that were used extensively by MITI in supporting the Japanese industries that it had selected for encouragement. In short, they are attractive means for use in carrying out the type of industrial policy that the formal analysis reported in this article seems to suggest.

3. **The Industries to Be Targeted for Encouragement.** Advocates of industrial policy frequently speak of the need for government to identify industries with promising prospects for growth as the proper subjects for measures that facilitate their establishment and take off. That is, promise of future expansion seems to be the attribute emphasized in these discussions as the criterion to be used in deciding which industries should be selected for government support and encouragement.

However, the analysis here indicates that there is another criterion that may be even more important than an industry's growth prospects. This is their retainability. That is, the model indicates that a scale diseconomies industry with no special technical features, is likely to be a poor target for support under industrial policy, even if it is likely to undergo rapid growth. As we have seen, such industries are not subject to "capture" by any one country, because by their nature one can expect many countries to be able to operate successfully and simultaneously in such a market. Moreover, a country's expansion in such a field is unlikely to represent more than a temporary gain unless comparative advantage has undergone a more durable modification making for such a change, in which case the need for any government intervention to achieve the expansion becomes questionable in the first place. In the absence of an accompanying change in the pattern of comparative advantage, gains in a scale-diseconomies industry achieved

through government subsidy, protectionist policy or other means will last only as long as the government support continues and as long as other nations do not retaliate. The moment public sector intervention ceases other nations will begin gradually to recoup their share of the markets at issue, because with diminishing returns such gradual expansion will quickly contribute to profits.

In contrast, once a retainable industry is gained, it is not easily recaptured by a rival economy. Thus, such an industry is worth fighting for, because its acquisition is likely to yield benefits for a considerable period. More than that. The governmental intervention required for the purpose is only temporary.

A second characteristic of an industry appropriate for encouragement by public policy is high elasticity of demand. An industry whose demand is elastic can expand its output without any substantial cut in price ensuing. Thus, growth in its volume of production is more likely to prove profitable the higher its demand elasticity.

Of course, growth prospects continue to be pertinent. There is little to be gained in acquiring an industry the demand for whose products throughout the world is eroding. Support of such dying industries never seems to make much sense, even if their technology is characterized by substantial economies of scale. And among industries with scale economies one that promises to expand more rapidly than the others is certainly a correspondingly more attractive target for industrial policy support.[19]

---

[19] Postwar Japanese industrial policy seems to have approximated some of the policy guidelines suggested in the text. A substantial proportion of the industries selected for encouragement seem to be characterized by scale economies. More of them seem to be N-Retainable (consumer electronics, film, computers). In addition, a major instrument used by the Japanese government to encourage these industries is reduction of the effective cost of capital. As already noted, incentives for banks to make funds available on easy terms to the firms in question are reported to have been a common recourse. This, as has been suggested in the text, is likely to minimize any consequent distortion of the market forces and any resulting and wasteful increase in social costs.

4. Relation of Industrial Economies and LDCs. The analysis suggests, finally, that the superiority in the economic position of the industrial countries over the LDCs may have gone too far, even from the point of view of the selfish interests of the former. It may well be that the relationship between the two country groups that is entailed in current equilibria has brought the industrial countries to the zone in which the utility boundaries of the industrial countries, like those of the LDCs, are negatively sloping. In that case, the mutual interests of the two groups of countries can be served by coordinated measures that are calculated to transfer some industries with scale economies to the developing countries. Despite the mutual benefits, such a program is likely to run into serious political obstacles in the developed nations. That is a subject on which a few remarks will be offered in the next and final section.

VIII. Concluding Comment: On the Political Problems.

Our analysis has suggested that rationality may call for substantial reorientation of policy in directions not generally recommended by earlier incarnations of trade theory. However, it is essential to take several political considerations into account before undertaking any attempt to achieve such policy modifications. There are at least two reasons. First, political circumstances can sometimes constitute a difficult hurdle for the adoption of rational policy, one which can be surmounted only by careful framing of the proposals at issue and by the utmost care about the manner in which they are presented. Second, even when political barriers can be overcome, the process can transform a desirable proposal beyond recognition, turning it into a caricature of what was intended, one that not only undoes the benefits of the original program but actually threatens serious damage to the public interest.

A clear case in point is that with which the preceding section concluded -- the difficulty of arranging for an industrialized country to give up one of its industries voluntarily to an LDC, beneficial though it may be for both in the long run. The difficulty, of course, is that it will not be easy to persuade an industrialized country to give up any of its industries, particularly because of the understandable protests by the industry's employees that any such attempt will predictably elicit. Thus, it appears politically impossible for any of the developed countries to make the first move despite the prospective long run benefits to its citizens. Some LDCs are, perhaps, more likely to be able to improve their own lot by their own actions. The examples

of Japan, Singapore, Taiwan and Hong Kong show that effective government intervention is possible in these highly disciplined societies. However, there are numerous failures that can be pointed to along with these successes. The efforts of Brazil appear to be a case in point.

Within an industrialized economy, and particularly in the United states, the problems are likely to be no less formidable. Indeed, the problems have repeatedly been pointed out by the opponents of industrial policy. There are at least two difficulties that are raised repeatedly: the lack of ability (or of demonstrated success) by government agencies in picking the "winner industries" that are to be granted assistance by the public sector, and the political pressures that are likely to force them to select the wrong industries and the wrong supportive measures even if the government agencies have all the information they need to do better than that.

The first of these difficulties, the difficulty of choosing, has, even in its present form, probably been overestimated. Governments around the world pick winners and losers, and *then try to make their choice come out right*. Their successes and failures probably have as much to do with their economy's ability to *make* the chosen industry into a winner as with their ability to "pick" one. In addition, as argued in this paper, a "winner" is likely to be characterized above all by its more-observable retainability, not by its more nebulous future growth prospects. Also, in the domain of our analysis the notion that the unaided private capital market is likely to do a far better job is something of a leap of faith. For in the domain that we have been discussing, the changes required to achieve durable improvement are not attained by small incremental modifications of the current state of affairs. Rather, as has repeatedly been emphasized here, what is typically called for is a discontinuous change, one that is likely to take the economy well beyond the range of experience of the business community. Thus, both business and government are likely to be proceeding on the basis of little reliable information in the arena under discussion. Either group is apt to need all the help it can get if it is to select moves with any reasonable prospects of success. Perhaps the computational procedures that underlie the theoretical analysis that this article has described may even be able to offer some help in the process.

The second of these caveats has, for example, recently been called to our attention once again by Alan Blinder [1992 pp. 27-28] who writes: "In Japan, industrial policy was and is run by a cadre of intelligent, respected, and powerful technocrats largely insulated from political

interference and acting in the national interest. The United States, I am afraid, is too democratic for that. Political considerations would quickly overwhelm economic merits; industrial policy would more closely resemble life support for dying industries than incubation of emerging ones." To this he might well have added that political pressures are all too likely to push our government toward increased use of protectionist means, like tariffs or, even worse, devices such as "voluntary" quotas upon exports to the United States, devices that present monopoly power as a gift to the countries with which the U.S. trades and with whose aid those countries can raise the prices and the revenues the American public pays to them. Thus, nothing said here is intended to minimize the difficulties likely to be encountered by rational industrial policy in the United States. Indeed it may point to the *necessity* of strengthening the U.S. Government's ability to understand and work with industry before any coherent industrial policy is attempted.

REFERENCES

Baumol, William J. and Gomory, Ralph E., "On Efficiency and Comparative Advantage in Trade Equilibria Under Scale Economies, 1992, Forthcoming.

Blinder, Alan S., "Should the Former Socialist Economists Look East or West for a Model?" Talk delivered to the International Economic Association, Moscow, August 28, 1992, Princeton University (unpublished).

Chipman, John, "A Survey of the Theory of International Trade: Part 2, The Neo-Classical Theory," **Econometrica**, **33**, October 1965, 685-760.

Ethier, Wilfred J., "Internationally Decreasing Costs and World Trade," **Journal of International Economics**, **9**, 1979, 1-24.

_____, "Decreasing Costs in International Trade and Frank Graham's Argument for Protection," **Econometrica**, **50**, September 1982, 1243-1268.

Gomory, Ralph E., "A Ricardo Model with Economies of Scale," **Economic Research Reports**, New York University, C.V. Starr Center for Applied Economics, (revised version) RR# 92-29, June 1992.

_____ and Baumol, William J., "Scale Economies, Regions of Multiple Trade Equilibria, and the Gains from Acquisition of Industries," **Economic Research Reports**, New York University, C.V. Starr Center for Applied Economics, RR# 92-10, March 1992.

Grossman, G.M. and Helpman, Elhanan, **Innovation and Growth in the Global Economy**, Cambridge, Mass.: MIT Press, 1991.

Helpman, Elhanan and Krugman, P.R., **Market Structure and Foreign Trade**, Cambridge, Mass.: MIT Press, 1985.

Kemp, M.C., **The Pure Theory of International Trade**, Englewood Cliffs, N.J.: Prentice Hall, 1969.

Krugman, Paul R., "History Versus Expectations," **Quarterly Journal of Economics, 106**, May 1991, 651-667.

Matthews, R.C.O., "Reciprocal Demand and Increasing Returns," **The Review of Economic Studies, 17**, 1949/50, 149-158.

Meade, James E., **A Geometry of International Trade**, London: George Allen and Unwin, 1952.

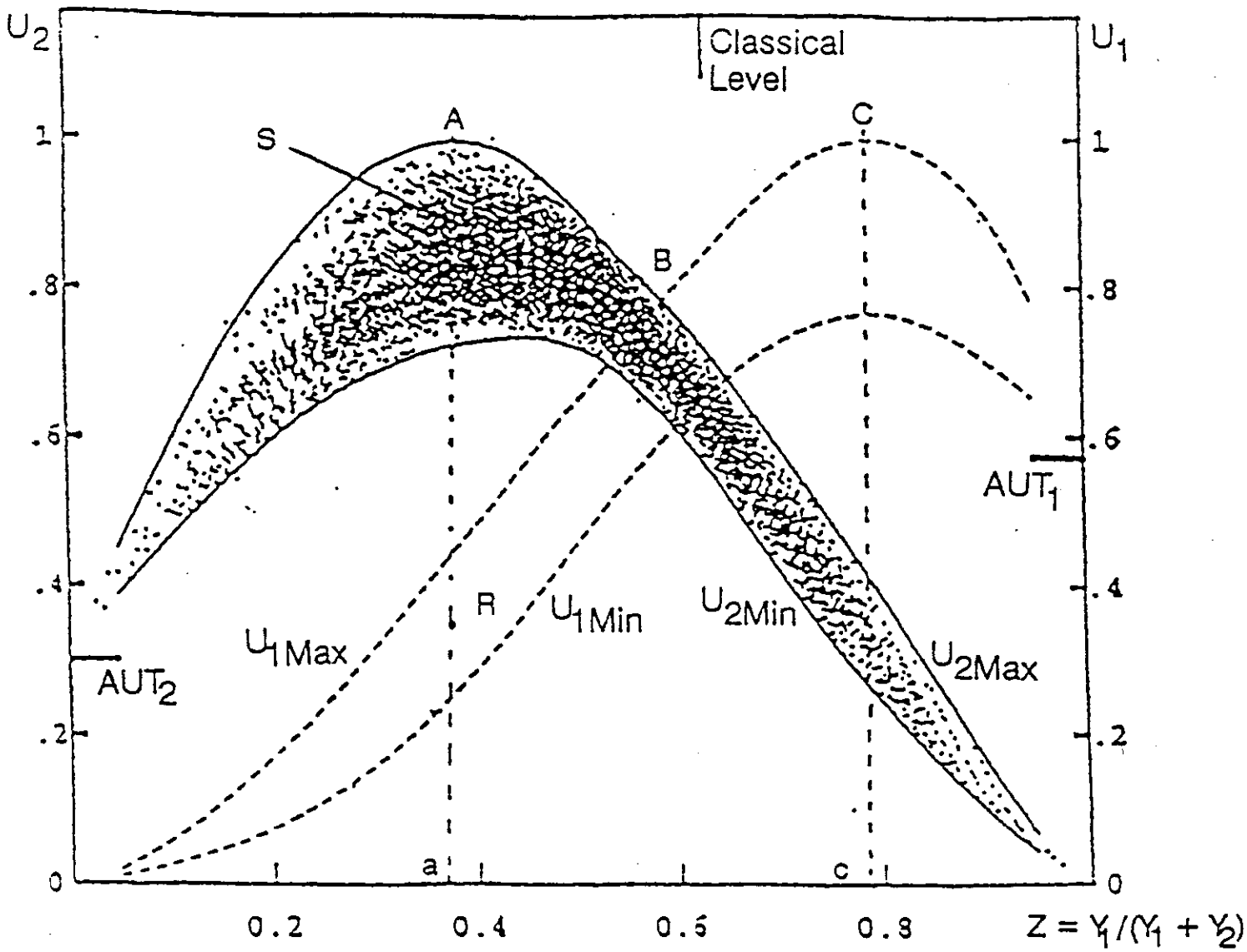
U2
U1
Classical
Level
S
A
B
C
R
a
c
AUT2
AUT1
U1Max
U1Min
U2Min
U2Max
1
.8
.6
.4
.2
0
0.2
0.4
0.6
0.8
Z = Y1/(Y1 + Y2)

Figure 1

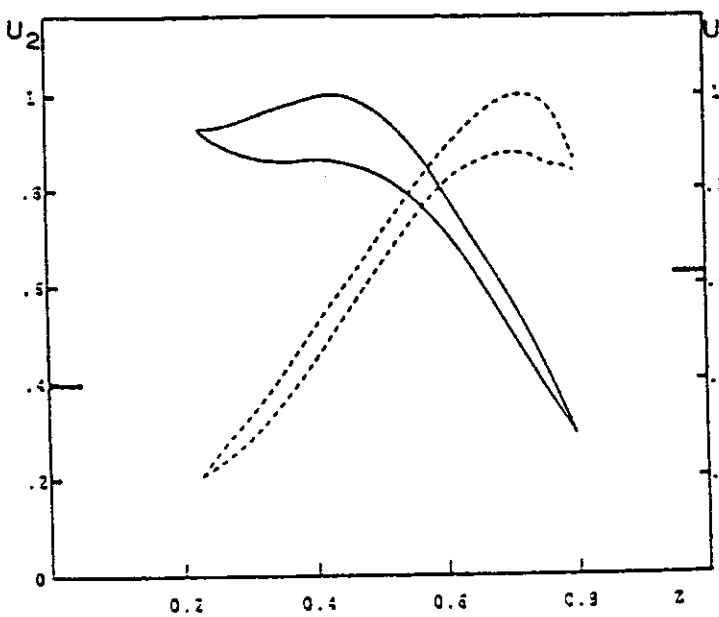

Figure 2A
25% Diminishing Returns
Industries

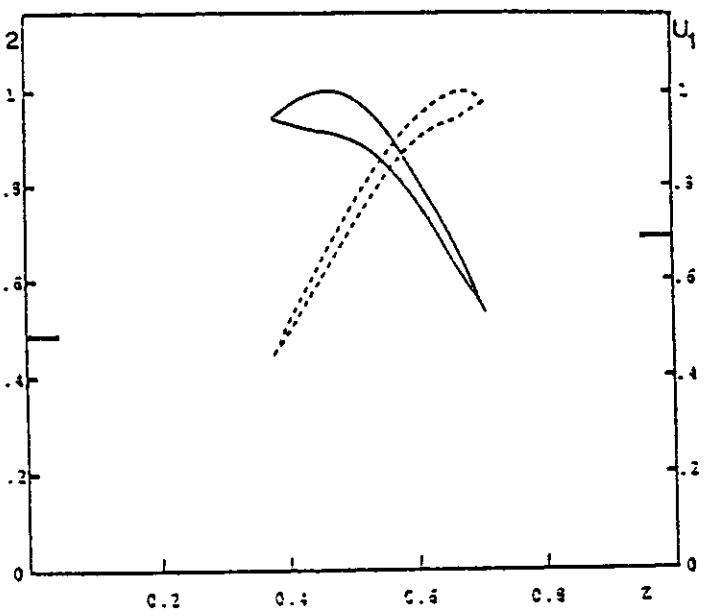

Figure 2B
50% Diminishing Returns
Industries

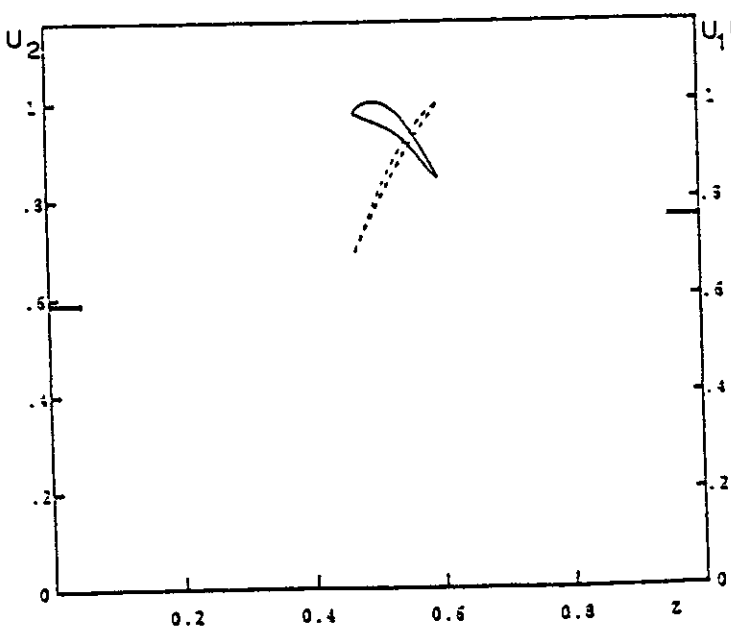

Figure 2C
75% Diminishing Returns
Industries

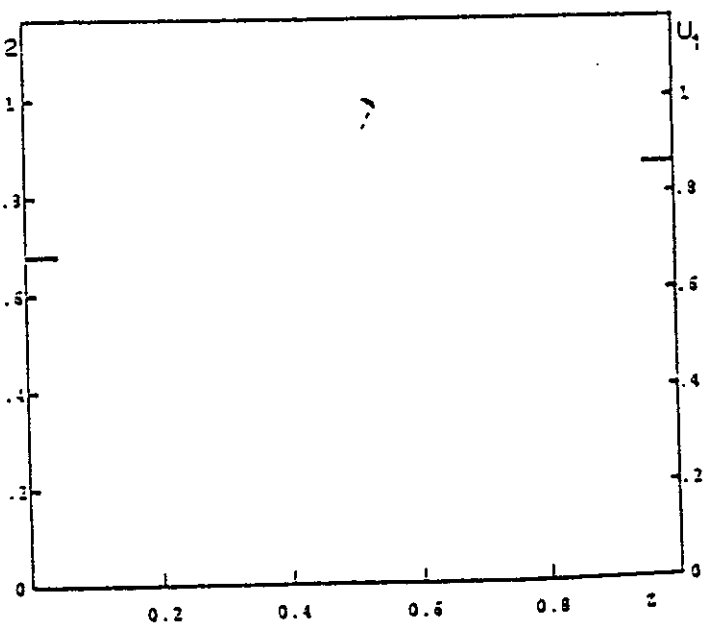

Figure 2D
95% Diminishing Returns
Industries

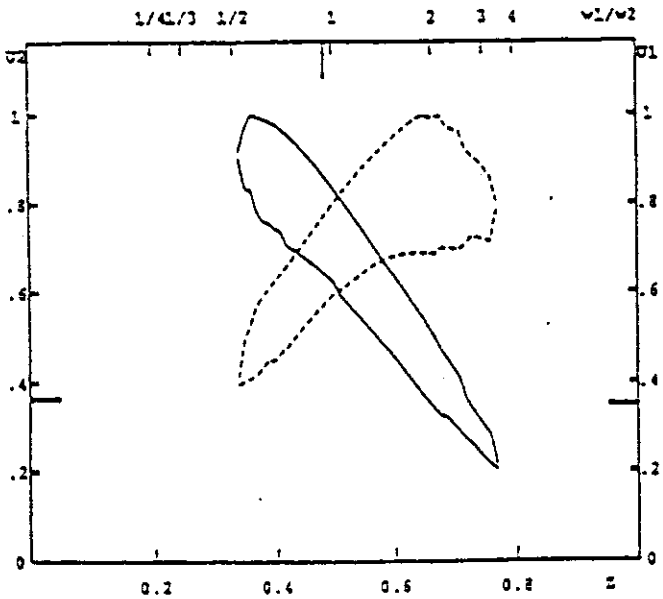

Figure 3A
$V_1 = 0.20$, $V_2 = 0.25$

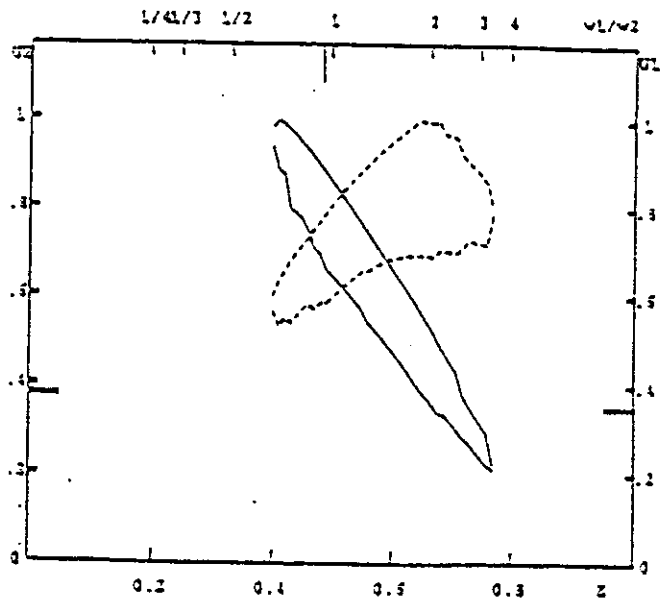

Figure 3B
$V_1 = 0.40$, $V_2 = 0.05$

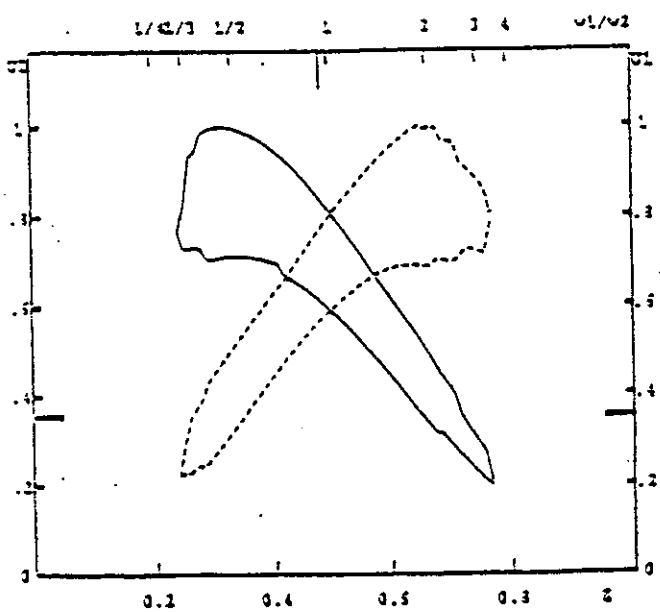

Figure 3C
$V_1 = 0.05$, $V_2 = 0.05$

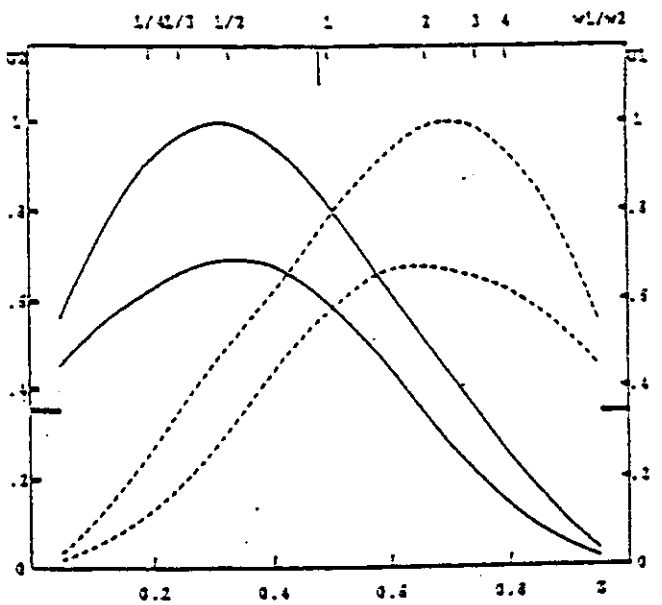

Figure 3D
$V_1 = 0$, $V_2 = 0$