A REFORMULATION OF THE INFANT INDUSTRY ARGUMENT

Tōru YANAGIHARA

I. INTRODUCTION

THE INFANT industry argument is widely accepted as a justifiable exception to the doctrine of free trade. For many economists, it represents the only case where a deviation from free trade is theoretically justifiable. The thrust of the infant industry argument is that temporary protection, even though a deviation from free trade, could lead to a more efficient allocation of world resources in the future, the gain more than offsetting the initial loss in allocative efficiency. Thus stated, the infant industry argument is a case for policy intervention for the purpose of realizing dynamic comparative advantage that is latent but will not come forth through the free play of market forces. The argument has been subjected to close scrutiny by several generations of neoclassical economists and much conceptual clarification and theoretical articulation has been achieved despite some controversy as to the arguments's essential features.[1]

The aim of this paper is to survey various issues accompanying the infant industry argument and to summarize the argument by giving it formal expression in a general decision-theoretic formula. As we proceed we will try to make explicit positive theories of infant industry development that underlie policy prescriptions.

II. THE ESSENCE OF THE INFANT INDUSTRY ARGUMENT

A part of the controversy concerning the essence of the infant industry argument is purely semantic, but the controversy as a whole has helped clarify quite a few confusions in the thinking of both academicians and practitioners. There now seems to be a concensus, at least among neoclassical economists, that technological progress is the essential underlying factor that governs the process of industrial maturing. And yet the characteristics of technological progress have not always been made explicit, and this has constituted one source of confusion in the field. In preparation for our discussion of the issue, the following table will classify technological progress into four types according to source and agent.

Agent \ Source	Experience	Investment
Firm	Learning by management	R & D
Worker	Learning by worker	Training

[1] An excellent survey of the issues and controversies is found in Corden [3].

We distinguish costless and costly processes of technological progress with respect to the source of the progress. Here, learning from experience in production is conceptualized as a by-product of the production of goods and is regarded as being obtained without incurring any extra cost. In contrast, there are instances of conscious investment decision to acquire technological progress. These investment decisions are separate, at least conceptually, from current operational decisions. The distinction is, of course, not entirely free from confusion. Since production is one source of technological progress, it is conceivable that the firm decides upon the level of production taking the long-term effect into consideration. In this case, the firm's investment in technological progress takes the form of extra costs in current product beyond the current-profit maximization level. The distinction between "experience" and "investment" in the table refers to the technical mechanism of technological progress and not to the nature of economic decision-making. Another point concerning this distinction between costless and costly learning was raised by Robert E. Baldwin. He claims that learning through experience involves direct costs for the firm because "it will be necessary for management to devote resources to analyzing previous performance before evaluating new productive practices" [1, p. 299]. This is an empirical issue to be settled through careful case studies but it is very possible that Baldwin overstates his case. For our analytical purposes, however, attendant problems can be skirted by distinguishing the costly part of learning through experience from the costless part and regarding the former as part of R & D. This distinction can be easily incorporated into a general model of knowledge creation in which learning effects and R & D interact with each other.

Now we turn to the distinction between the firm and the worker as agents of technological progress. This distinction is important because knowledge acquired by the firm has the nature of a public good in the sense that its use by one firm does not preclude its use by other firms.[2] On the other hand, skill acquired by the worker is a private good in the sense that it can be employed by only one firm at a time. The opportunity cost of the use by one firm, measured as the amount of production foregone in the firm where technological progress originally took place, is zero in the case of knowledge itself and is equal to the marginal product in the case of the skilled worker. It should be noted that the above-mentioned distinction is between disembodied and embodied technological progress; it is not a distinction between general and firm-specific knowledge or skill [2]. This latter distinction adds another dimension to the table above. A firm's knowledge and a worker's skill are both either general or firm-specific (or somewhere in between).

Our formulation of the infant industry argument in the next section of this paper is concerned only with knowledge creation within the firm. The formal model presented there can be interpreted as including skill formation, but with some loss of theoretical rigor. There are a couple of justifications for the decision

2 The use of knowledge by other firms could affect the originating firm financially. There is rivalry in use in this sense and the firm will have incentive to take advantage of its possession of knowledge. There seems to have been some confusion in terminology in the discussion of the public-good nature of knowledge.

to limit the formulation to intrafirm knowledge creation besides the apparent one of ease in theoretical formulation. Gary S. Becker has shown that under perfect labor and capital market conditions, general training is financed by the worker, not by the firm [2]. Therefore its effects are external to the firm's decision-making. Specific training can be financed either by the worker or by the firm, but in any case will not pose a serious investment problem to the firm since, under perfect market conditions, the firm will finance specific training only when it is assured of a sufficiently high rate of return. In the case of some firms, however, knowledge creation has to be financed even if it is costly and the return to the firm is subject to a number of factors beyond its control. This provides possible cases for government intervention. Admittedly, labor and capital markets are not as perfect as required for the theoretical conclusion to hold in the case of skill formation, but this is a question of market imperfections and is assumed away in our theoretical model. A more realistic, albeit speculative, justification for our decision on the essence of the infant industry argument is the view that the growth of firms is the key element in industrial development. The firm is an entity in which capital, both tangible and intangible, is accumulated over time and whose efficiency and competitiveness is largely dependent on its intangible capital stock (here called knowledge) [12]. The time framework of the infant industry argument is even longer-run that the long-run in textbook accounts of firm behavior. The only firm-specific factor then is intangible capital stock, physical capital and labor being purchased in the market.

III. FIRM BEHAVIOR AND MARKET EQUILIBRIUM

The infant industry argument is concerned with two types of decisional problems: the first, whether an industry should be temporarily protected; and the second, supposing the answer to the first question to be affirmative, what policy measures should be adopted. The infant industry argument was originally put forward as a case for tariff protection, but it has since been generalized as a question of optimal policy intervention.[3] Here the term "protection" is to mean "policy intervention to foster the infant industry" and is not limited to "tariff protection." The first problem has been approached by postulating conditions under which protection is justified. The search for necessary and/or sufficient conditions for justifiable infant industry protection has led to two clarified understandings: that protection is required to cope with the problem of appropriability which an otherwise perfect market mechanism cannot resolve; and that there are two possible sources behind inappropriability situations: dynamic internal economies and dynamic external economies.[4] The term "inappropriability" is used broadly here to denote any excess of social benefits over private returns and thus encompasses both pecuniary and technical externalities.

Our theoretical model of the infant industry argument is built upon Negishi's

[3] See Johnson [5] for a unified approach to various cases.

[4] Negishi [10]. As Negishi points out, as is usual in the field of international economics and welfare economics, these apparently "new" ideas are presented and illustrated in Meade [8].

formulation of dynamic internal economies due to "the indivisibility which is inherent in the learning process of infant industries" [10, p. 58]. Our analytical framework is one of partial equilibrium and is more practically-oriented than a theoretically complete general equilibrium analysis of which the main part of Negishi's theoretical argument is an example. Also, unlike Negishi, we impose the small country assumption both for the present and the future. Thus we have only to be concerned with welfare for a particular country. Our's is therefore only a partial view of dynamic comparative advantage, but again our construct and the questions posed therein seem to have a closer relationship to actual policy formulations in infant industry protection. We will discuss dynamic external economies only briefly as a modification of the basic model, because the essential policy implications of dynamic external economies lie in externalities and not in their dynamic aspect. They have dynamic implications in that they affect investment decisions, but policy intervention would also be called for in order to correct externalities in cases of static external economies. And much has been already said about externalities.

A. *Model of Firm Behavior*

To the best of my knowledge there has been no previous attempt to formally set down assumptions on aspects of firm behavior that underlie the policy discussions of infant industry protection. First we will develop a model of firm behavior[5] under the assumption that there are no dynamic externalities; that is, that knowledge created remains within the firm. This presumes that all knowledge is firm-specific or the cost of transmission and absorption is higher than the cost of knowledge creation. We maintain this assumption despite the simplifying assumption we adopt in our proofs that the firms are otherwise identical.

As to market structure, we consider the polar cases of perfect competition and monopoly in the product market. In the context of the infant industry argument we assume that either there are N potential entrants (N large enough and each firm small enough) or that there is only one possible entrant to the industry. In each case, the initial market structure is maintained over time. Also it should be remembered that we maintain the small country assumption throughout. With respect to the factor market, the assumption is that the industry (if started) is initially too small to affect the factor price but that it will eventually grow large enough to affect the factor price whether under perfect competition or under monopoly. The monopolist then behaves as a monopsonist as well.

We assume the objective function of the firm to be maximization of the present value of the cash flows, with maximization of current profit as a special case. The special case is supposed to capture the behavior of myopic or technologically stagnant firms, but their behavior is rational insofar as the discount rate they adopt does not reflect an "irrational" fear of the future. The competitive firm is assumed to form a point estimate of future prices of the product and factors

[5] Modeling in this section is patterned after, and is a synthesis of, Nordhaus [11] and Rosen [13]. Corden [3] expounds the basic idea underlying the formulation here.

and treat them as certain and as exogenously given.[6] The monopolist estimates the demand schedule for his product and the supply schedule of the factors of production in solving his maximization problem. The production function of the firm is as follows:

$$Q=F(L,\ Z)=Zf(L). \tag{1}$$

$$f'(L)>0,\ f''(L)<0,\ f(0)=0.$$

In (1), L is the physical factor of production called "labor," which can be regarded as a composite of inputs combined in fixed proportions (under the assumption of weak separability). Z is the intangible factor of production called "knowledge." The particular features of the production function should be noted here. First, as one way to indicate the indivisibility of "knowledge" in use, Z enters the production function (1) in the multiplicative form. As a consequence, the marginal product of L is $Zf'(L)$, i.e., it is proportional to Z for a given level of L. Secondly, the marginal product of Z is equal to $f(L)$ and increases with L. This means that the larger the firm, the larger contribution a given addition to "knowledge" makes. Also note that the marginal product of Z is constant for any given level of L. This is partly definitional as is made clear below. The production function (1) shows an increasing return to scale with respect to L and Z so far as $f'(L)>0$, as is assumed here. This, however, poses no problem to the competitive solution since at any moment Z is predetermined for each firm and cannot be altered instantaneously. The usual concept of returns to scale corresponds to $f''(L)$ in our model. Thus the negativity of the second derivative is not as innocuous an assumption as its formal familiarity leads us to believe. In any time period we are dealing with the long-run relations, thus $f''(L)$ corresponds to diminishing returns to scale. One plausible reason for diminishing returns in this context is diseconomies of scale due to the fixity of managerial and organizational capabilities of the firm. In fact, Z in our model can be taken to represent just that; diminishing returns always set in, but a higher Z makes that happen at a higher level of productivity, enabling the firm to expand its size over time.

"Knowledge" is accumulated through learning and through R & D. The level of "knowledge" at time $(t+1)$ is determined by its level at time t, the amount of learning hypothesized to be a function of factor input at t, L_t, and the amount of R & D at time t, R_t:

$$Z_{t+1}=A(Z_t,\ L_t,\ R_t). \tag{2}$$

This formulation is sufficiently general to incorporate all sorts of returns, substitutabilities or complementarities. We can add time itself, calendar time and/or some sort of vintage, if we wish more generality. A note on the specification of the learning effect: conventional practices that relate learning to output, Q_t, are compatible with the "knowledge accumulation function" (2), since it is a function of Z_t and L_t as is specified by the production function (1). We will assume

6 Alternatively, the firm can be assumed to maximize the expected profit or a risk-discounted profit. The crucial assumption here is that the firm behaves as a price-taker.

sufficiently diminishing returns to scale with respect to L_t and R_t, so as to guarantee a stable equilibrium for the firm. As is clear in the specification of the production function (1), Z is defined and measured in efficiency units, so to speak. Thus function (2) reflects not only the creation but the implementation of new knowledge and is affected by the cost of growth as well as the returns to learning and R & D.[7] The inclusion of Z_t as a determinant of Z_{t+1} is straightforward if all or a part of old knowledge remains effective. The specification is meaningful even when that is not the case if old knowledge contributes to the creation of new knowledge, as seems likely.

In what follows we will consider a two-period model of firm behavior. Extension to further time periods is straightforward. Each period corresponds to the long run in the textbook sense. In period 0 the firm decides whether to enter the industry currently completely dominated by imports in view of the long-term profit potentials. The firm's objective function is as follows:

$$\max_{L_0, L_1, R} \{[p_0Q_0 - w_0L_0 - sR] + [p_1Q_1 - w_1L_1](1+r)^{-1}\}, \tag{3}$$

where p, w, and s are prices of Q, L, and R, respectively, and the subscripts 0 and 1 denote time period. The constraints are the production function (1) and the "knowledge accumulation function" (2), namely,

$$Q_t = Z_t f(L_t), \quad (t=0,\ 1),$$

$$Z_1 = A(Z_0,\ L_0,\ R).$$

It is assumed that the firm attains maturity in period 1 and R & D is no longer conducted. The solution for the competitive firm is given by the following set of simultaneous equations (assuming an inner solution):

$$[p_0Z_0f'(L_0) - w_0] + \left[p_1\frac{\partial A}{\partial L_0}f(L_1)\right](1+r)^{-1} = 0, \tag{4a}$$

$$-s + \left[p_1\frac{\partial A}{\partial R}f(L_1)\right](1+r)^{-1} = 0, \tag{4b}$$

$$[p_1Z_1f'(L_1) - w_1](1+r)^{-1} = 0. \tag{4c}$$

We have assumed that the production function and the "knowledge accumulation function" have such diminishing returns properties as will guarantee the second-order conditions for the maximization. Let us first examine (4b). It tells us that R & D is carried out up to the level where the marginal value product (properly discounted) equals the marginal cost. Roughly speaking, the marginal value product is proportional to $f(L_1)$, the size of the firm in the future. Turning now to (4a). We observe that the marginal value product of L_0 consists of two components, current and future. In other words, the current labor input is increased beyond the level where the current marginal value product equals the wage rate. Or put yet differently, the firm incurs an opportunity cost to "invest" in the learning process that contributes to knowledge accumulation. Equation (4c) is the straightforward static profit maximization condition that determines the

[7] See Penrose [12] for an elaborate discussion of the cost of growth.

size of the firm in period 1, with Z_1 predetermined by the levels of L_0 and R.

Now let us consider a number of special cases (corner solutions). If, for whatever reason, the discount rate, r, is prohibitively high, the firm's behavior will be myopic, and the firm will tend to be technologically stagnant. Such firms will continue to operate in period 1, however, if market conditions allow. Even if the discount rate is not that high, if the firm's prospects for knowledge accumulation are dim, either objectively or subjectively, the firm will appear to behave myopically. In such a case, whatever the underlying reasons, the firm will not enter the industry in period 0 if it cannot earn current profits, and it will not engage in R & D. Needless to say, if the firm is completely myopic, no R & D is performed unless its unit cost, s, is set to zero. If, to take another possibility, learning through experience does not contribute to knowledge accumulation significantly, the firm may engage solely in R & D in period 0 in preparation for entry into the industry in period 1. With this discussion of special cases behind us, we will henceforth only deal with inner solutions for the firm.

B. *Market Equilibrium*

Let us now consider the market equilibrium in periods 0 and 1. In period 0, it is assumed that domestic supply (from N firms altogether) does not completely supplant imports.[8] Hence the industry as a whole faces a perfectly elastic demand schedule. An assumption of infinite supply elasticities is made for the factor markets for L and R as well. The industry supply and demand schedules are obtained by simply adding up the respective schedules for the various firms (see Figure 1).

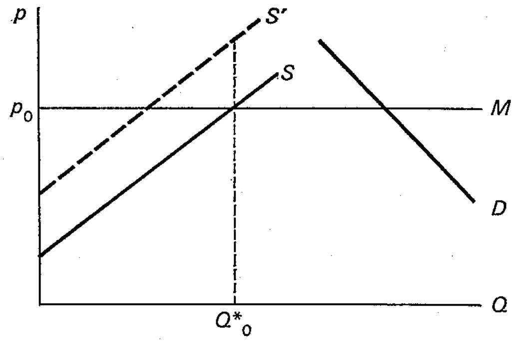
In the product market, market equilibrium is reached at Q^*_0, the point of intersection of the horizontal import supply schedule, M, and the domestic supply schedule, S, which is the net marginal cost curve for the industry—net in the sense that the (anticipated) marginal revenue in period 1 is subtracted from the current marginal cost, whose schedule is given as S'. In the labor market, we observe the dual expression of this divergence between S and S'. Labor market equilibrium is reached at L^*_0, the point of intersection of the horizontal supply schedule, S, and the "present-value marginal product" curve, D, which lies above the "current-value marginal product" curve, D'. Market equilibrium for the R & D market is determined by the same principle. In this case, the "present-value marginal product" consists solely of future revenue and D' coincides with the vertical axis for any positive value of s.

In period 1, we assume that domestic industry is large enough to affect the labor market, although we maintain the assumption that each of N firms in the industry is sufficiently small and behaves as a price-taker. A similar set of assumptions holds with regard to the product market, except that equilibrium may be attained at the horizontal portion of the demand schedule for the domestic

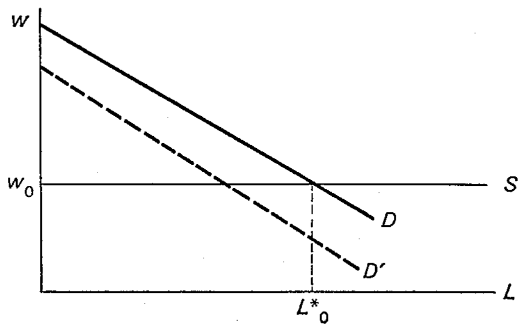
8 This assumption, along with the assumptions with regard to the factor markets that follow, is adopted to simplify the argument. Modifications for less than infinite elasticities are easily achieved along the lines indicated for period 1. It is assumed that government does not prohibit imports.

Fig. 1. Market Equilibriums in Period 0

a. Product market



b. Labor market



c. R & D market

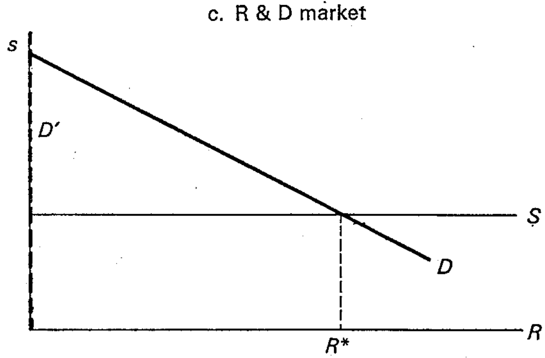


Fig. 2. Demand Schedule for Domestic Industry in Period 1

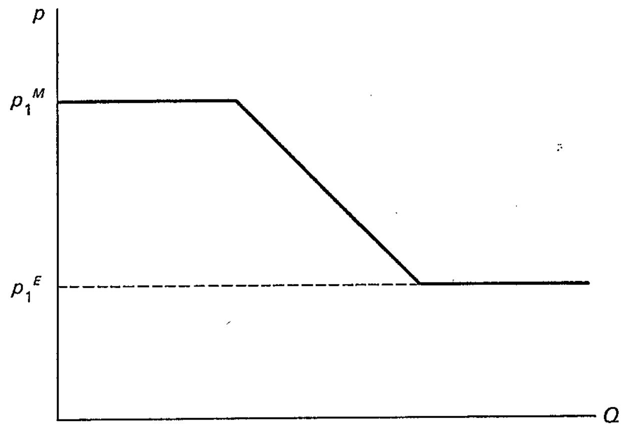
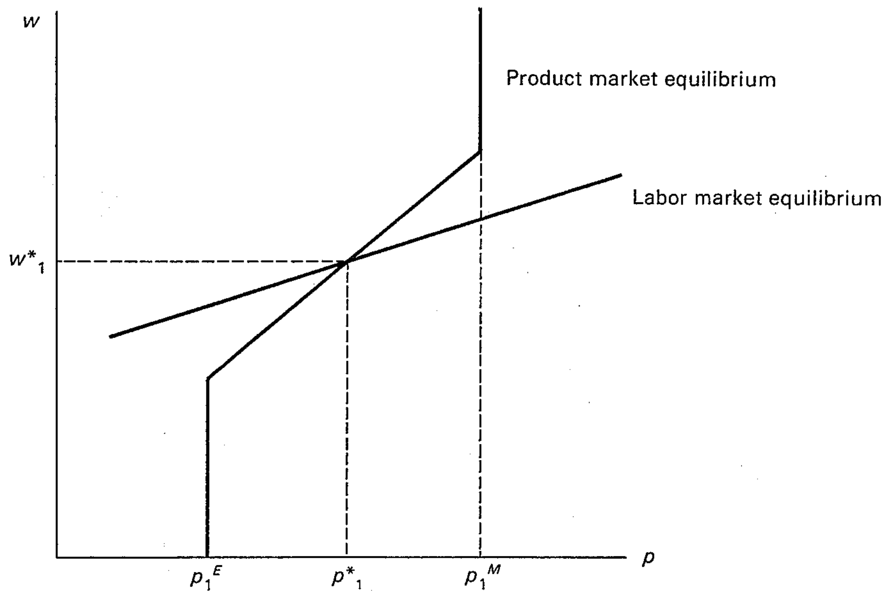


Fig. 3. Determination of Equilibrium Prices in Period 1



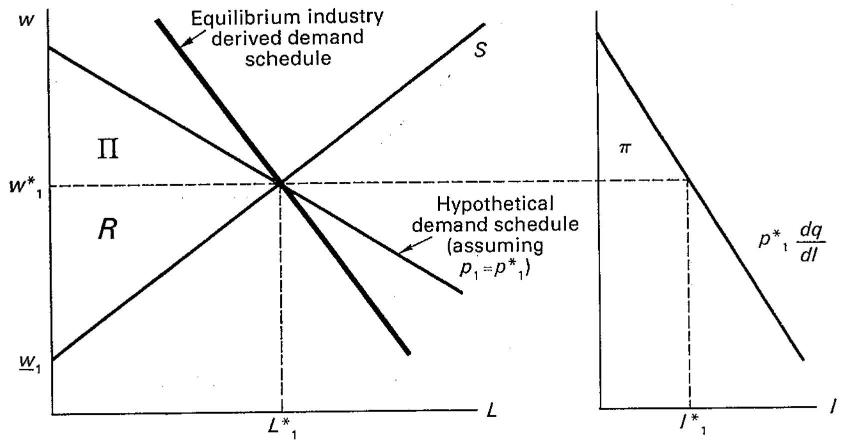
Note: See Appendix for technical notes.

industry (see Figure 2). There is no demand for domestic products above the supply price of imports, p_1^M, while at p_1^M and at the f.o.b. export price, p_1^E, there is an infinitely elastic demand schedule. Here we take up the case in which market equilibrium is reached at the downward sloping portion of the demand schedule; the domestic industry has attained self-sufficiency, but has not yet turned into an export industry. The case where market equilibrium is reached at one of the horizontal sections of the demand schedule can be considered as a special case of the general framework developed below.

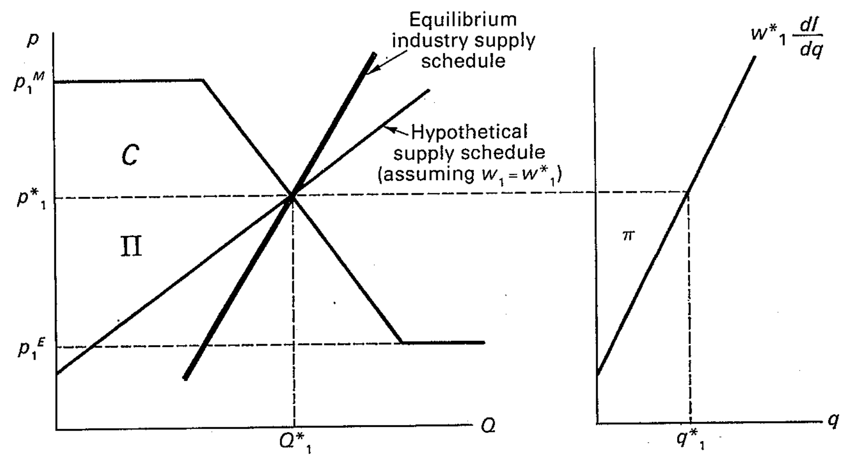
Market equilibriums in the product and labor markets are determined simul-

Fig. 4. Equilibrium in the Market and the Firm in Period 1 (Competitive Case)

a. Labor market

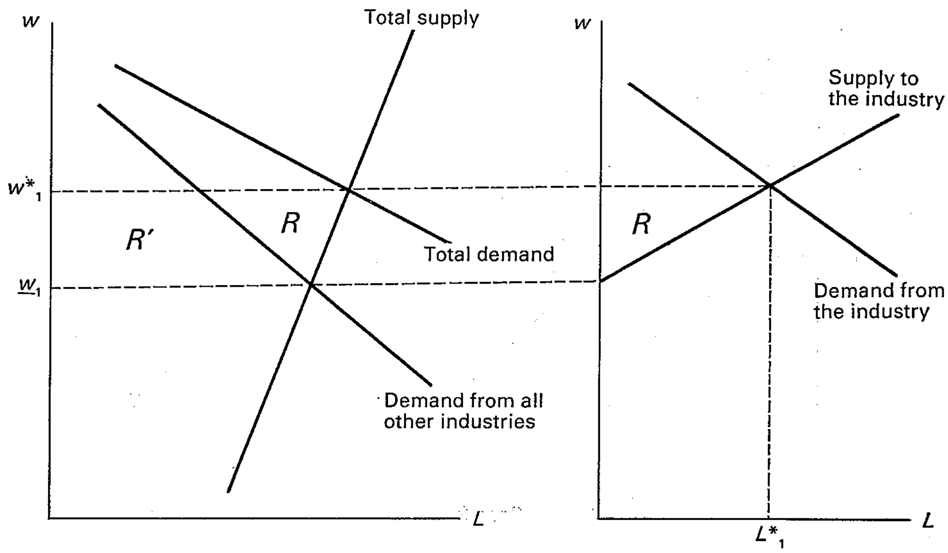


b. Product market



taneously in the following manner. A hypothetical industry supply schedule under the (false) assumption of a constant wage rate is obtained by adding up the supply schedules of the individual firms under the same (for each of them, correct) assumption. By varying the wage rate, w, we obtain a relationship between p and w that is associated with equilibrium in the product market (all this while keeping the demand schedule fixed). A similar procedure for the labor market (changing the roles of p and w and keeping the supply schedule fixed) gives a relationship between p and w that is associated with equilibrium in the labor market. The equilibriums in the two markets are simultaneously determined at values of p and w that satisfy both relationships (see Figure 3).

Fig. 5. Determination of Wage Rate and Increase in Rent Accruing to Labor



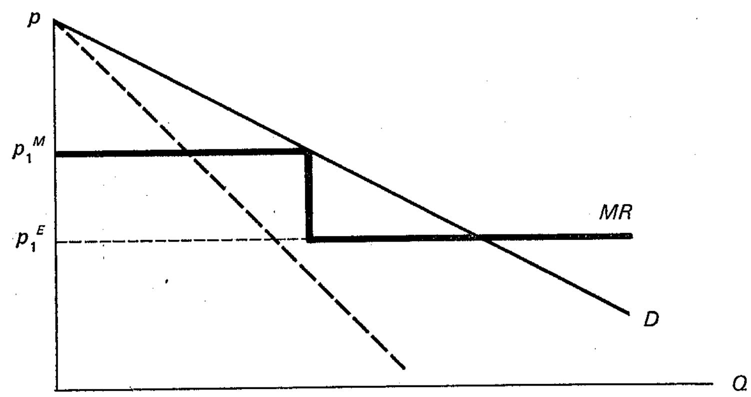
The profit, π, accruing to the industry under equilibrium conditions is given by the area formed by the horizontal line of height p^* and the hypothetical supply curve (under the assumption of a constant wage rate, w^*) in the diagram for the product market (Figure 4b). Or, alternatively, it is given as the area formed by the horizontal line of height w^* and the hypothetical derived demand curve (under the assumption of a constant price, p^*) in the diagram for the labor market (Figure 4a). The equilibrium industry supply schedule and the equilibrium industry derived demand schedule are shown by thick lines in Figure 4b and Figure 4a, respectively. (Their derivation is illustrated in Appendix.) The only reason to show them in the diagrams is to demonstrate that they should *not* be used to find the profits accruing to the industry.

In Figure 4a, total wage payments are equal to $w^*_1L^*_1$, of which the triangle marked R is rent to labor employed in the industry. Here it should be noted that the schedule of supply of labor to the industry depicted in Figure 4a is derived by subtracting the equilibrium demands for labor of all other industries from the overall supply schedule of labor. Thus we expect it to be rather elastic. But insofar as the demand from this industry causes the market wage rate to rise, the benefits extend to labor employed in other industries as well. Therefore the rent accruing to the entire labor ($R+R'$), must be measured along the total supply schedule, not just along the supply schedule to the industry in question (Figure 5).

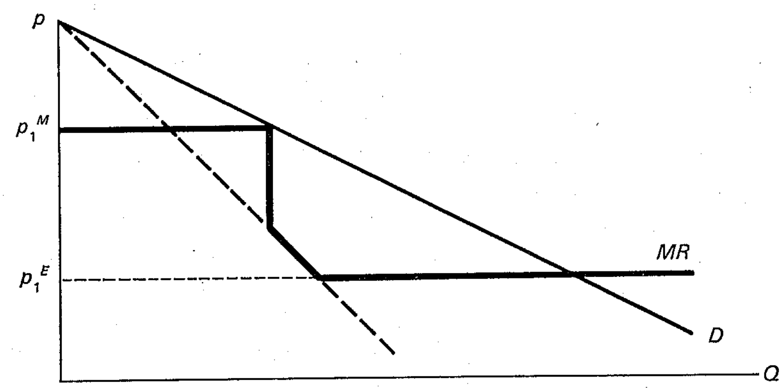
In Figure 4b the trapezoidal area marked C indicates the increase in consumers' surplus over what it would be if there was no domestic supply. The area will be nil if the domestic industry falls short of achieving self-sufficiency in period 1. If, on the other hand, the domestic industry grows to become an export industry, the trapezoid C will be expanded downward to have the height $(p_1^M - p_1^E)$.

Fig. 6. Marginal Revenue Curve for the Monopolist

Case 1



Case 2

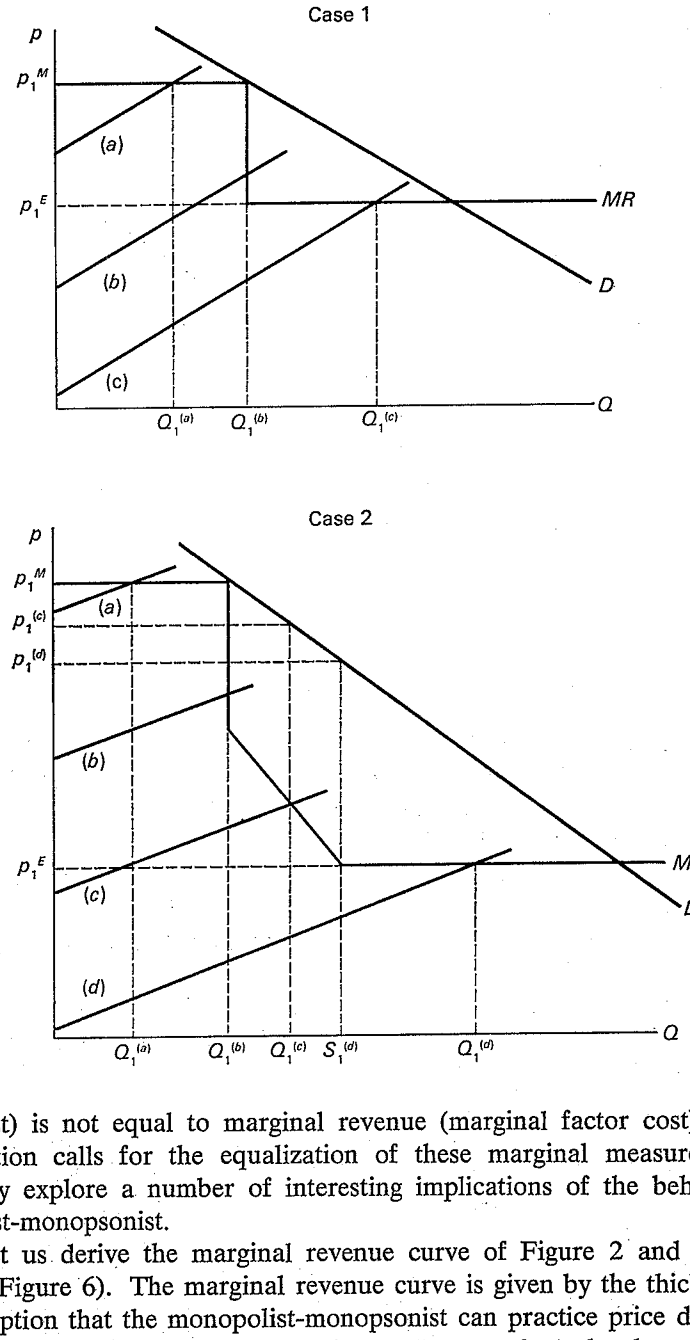


C. *Monopolist-Monopsonist Case*

Let us now turn to the opposite extreme and assume that there is only one potential entrant to the industry in period 0. The monopolist-monopsonist[9] estimates the product demand and the labor supply schedules facing him in period 1 and tries to maximize his profits by equating the marginal revenue and the marginal cost (or, equivalently, by setting the marginal revenue product equal to the marginal factor cost). Our theoretical formulation of firm behavior holds in this case as well simply by replacing product price, p, by marginal revenue and wage rate by marginal factor cost in the first order conditions, (4a), (4b), and (4c). Under our assumptions as previously stated, the firm faces perfectly elastic product demand and labor supply schedules in period 0, so p_0 and w remain as they are in the trio of equation (4). In period 1, however, average revenue (average

[9] The term "monopsonist" should be qualified to mean those who are the sole purchasers with respect to the residual supply schedule facing them in the sense discussed previously.

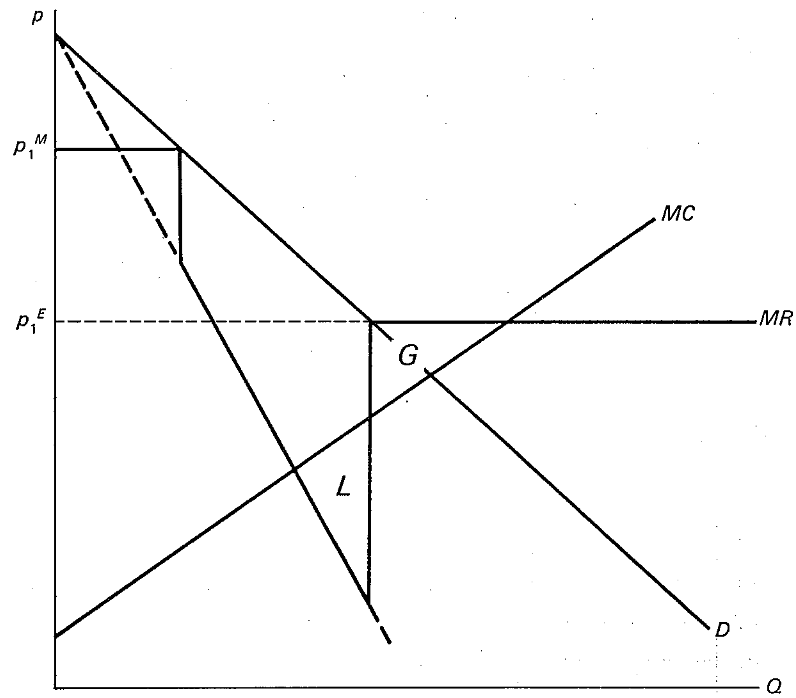
Fig. 7. Marginal Revenue Curve and Marginal Cost Curve for the Monopolist



factor cost) is not equal to marginal revenue (marginal factor cost), and profit maximization calls for the equalization of these marginal measures. We will graphically explore a number of interesting implications of the behavior of the monopolist-monopsonist.

First let us derive the marginal revenue curve of Figure 2 and reproduce it here (see Figure 6). The marginal revenue curve is given by the thick line, under the assumption that the monopolist-monopsonist can practice price discrimination between the domestic and export markets. We reproduce the demand and marginal revenue curves and draw in alternative positions of the marginal cost curve (see Figure 7).

Fig. 8. Marginal Revenue Curve without Price Discrimination



First we take up case 1. For all three alternative positions of the marginal cost curve, (*a*), (*b*), and (*c*), profits to the monopolist are represented by the area circumscribed by the marginal revenue curve, *MR*, the marginal cost curve, and the vertical axis. The domestic price is set at p_1^M whether part of the output is exported or not, and no excess of consumers' surplus is obtained compared to the no-domestic-supply situation. Rent accruing to total labor is found by integrating the total labor supply schedule from $\underline{w}$ to w^*_1 (Figure 5), as in the competitive case.

Now we turn to case 2. What is new in this case is the introduction of two alternative product prices. The domestic price is set at $p_1^{(c)}$ and $p_1^{(d)}$, respectively, and consumers' surpluses are generated. Everything else discussed for case 1 holds true with case 2.

Here let us drop the assumption of market discrimination and assume, instead, that the monopolist has to charge an identical price in both domestic and export markets (in that case the single price will be the export price, p_1^E, given exogenously).[10] Under this new assumption, the marginal revenue curve up to the level of output at which the domestic demand schedule intersects with the (horizontal) export demand schedule is the one associated with domestic demand (see Figure 8). In Figure 8, the monopolist will export if and only if the area marked *G*

[10] Harry G. Johnson seems to have adopted this assumption without explicitly mentioning it in his article [6, p. 72].

is larger than the area L; otherwise, he will only supply to the domestic market. Thus, compared with the monopolist's solution under the original assumption of price discrimination, he is now less likely to export and, when he does, he exports less, given a marginal cost curve; the domestic price, under that situation, will tend to be lower. Also note that when the monopolist actually exports under the condition of no price discrimination, the quantities supplied domestically and abroad are exactly the same as those supplied by a perfectly discriminating monopolist, or by competitive industry for that matter, although the nature of the industry marginal cost curve is different for competitive industry.

D. *Dynamic External Effects*

We will be very brief on externalities, taking up only one particular specification for our discussion. We maintain the production function, (1), but replace the "knowledge accumulation function" for firm i by the following:

$$Z^i_{t+1}=\sum_{j=1}^{N}\lambda_{ji}A^j(Z_t^j,\ L_t^j,\ R_t^j),\quad (i=1,\ \cdots,\ N). \tag{5}$$

In this formulation, part of the knowledge created in other firms is transferred and incorporated into the intangible capital stock of firm i. The firm's objective function is the same as before, i.e., (3), because this transfer of knowledge is assumed to be costless. Its decision on L_0, R, and L_1 now hinges on how it expects others to behave, however. A situation could arise in which every firm tries to be a free-loader and none takes the initiative. We will only try to be illustrative here and assume that externalities are symmetrical and that all firms are identical ex-ante and ex-post (we also assume that firms in our model operate on that assumption). These assumptions allow us (and the firms) to solve the maximization problem, (3), to find the optimum values L^*_0, R^*, L^*_1, and therefore Z^*_1. The first order conditions are formally, only slightly modified, the only changes being the insertion of λ_{ii} in front of $\partial A/\partial L_0$ in (4a) and in front of $\partial A/\partial R$ in (4b). Now dropping superscripts i and j and denoting λ_{ii} by λ and $\lambda_{ji}(j\neq i)$ by μ, the planned values, L^*_0, R^*, and Z^*_1 should satisfy the following relationship:

$$Z^*_1=[\lambda+(N-1)\mu]A(Z_0,\ L^*_0,\ R^*_0). \tag{6}$$

Once Z_1 is determined, for the discussion of period 1, we go back to the original no-externality case.

IV. SOCIAL COST-BENEFIT ANALYSIS OF INFANT INDUSTRY PROTECTION

In his oft-quoted article on the infant industry argument, Herbert G. Grubel gave an analytical expression to the "Bastable test" and provided a cost-benefit analysis framework for the argument [4]. He adopted two alternative assumptions concerning the industry supply curve in the long run, the infinitely elastic supply case and the upward-sloping supply curve case.

The industry supply curve is assumed to shift with time in either case. For the horizontal supply curve case, the imposition of a prohibitive tariff is assumed to be in effect so far as domestic cost exceeds the world price. Under these conditions, the social cost of protection is equal to the reduction in consumers' surplus over the period protection is needed. The social benefit of protection, on the other hand, is equal to the increase in consumers' surplus when the domestic supply price becomes lower than the world price. He runs into a difficulty here, however, because this country, under these assumptions, would become the single supplier in the world. Here he switches to the alternative assumption of the rising long-run supply curve (he is not explicit as to what assumptions lie behind this upward-sloping curve), under which the social cost of tariff protection is comprised of the deadweight losses in consumption and in production. The social gains of tariff protection of the industry are realized as increases in consumers' and producers' surpluses once the domestic industry reaches maturity. Assuming that the prices involved reflect social opportunity costs of resources, he plugs these costs and benefits into the discounted present value formula. Here we will try to further develop the line of theoretical argument initiated by Grubel.

A. *A General Formulation of the Governmental Decision-Making Problem*

Grubel concluded his article on the infant industry argument for tariff protection on a negative note, pointing to the shortcomings of the argument on two accounts, "first the lack of information for empirical identification of industries and second the availability of more efficient alternative policies" [4, p. 338]. His first criticism could be turned into an across-the-board negation of the value of rational policy-oriented approaches to the choice of industrial structure. (It is not quite clear whether this verdict of his was meant to apply to social cost-benefit analysis in general or to a cost-benefit approach to the infant industry argument in particular—more on this later.) His second criticism is concerned with the choice of policy tools, tariff vs. subsidy. He recommends subsidy over tariff as a means of infant industry protection, arguing that subsidy spares the country distortions in consumption caused by tariff during the period of protection. But if Grubel were to be as empirically realistic on this second point as on the first, he would certainly note distortions caused by the tax-raising needed to underwrite the necessary subsidies. In this respect his proposal of a self-financing tariff-cum-subsidy scheme [4, p. 340] has no theoretical claim to superiority in static allocative efficiency. Tariff itself is a particular kind of self-financing subsidy scheme (actually it over-finances from the viewpoint of protective effects), one that subsidizes in the form of higher price. The only general principle we can turn to here is that subsidy should be financed by the method that creates the least distortions in the economy [3, pp. 43–45]. In what follows we will assume that such a method of tax collection does not involve tariff on the industry to which subsidy is to be provided. We will discuss Grubel's proposal again later in a somewhat more down-to-earth context, but for now we will continue to be idealistic concerning the relevance of economics to the solution of economic problems of major proportions, such as social cost-benefit analysis of infant industry protection.

We will now combine two types of decisional problem involved in the infant industry argument: choice of industry and choice of policy. A general formulation of governmental decision-making is as follows (we still stick to a two-period world; extension to an N-period model is straightforward):

$$\max_{P_0} [NB_0(P_0)+NB_1(P_0)(1+rs)^{-1}]. \quad (7)$$

NB stands for net benefit (to society), rs is the social rate of discount, and the subscripts 0 and 1 refer to periods. Formulation (7) tells us that government should choose a policy in period 0 that will lead to maximization of the present value of social net benefits in periods 0 and 1. Here we have implicitly assumed that laissez faire prevails in period 1, to emphasize the temporary nature of infant industry protection.[11] Here government's problem is formulated as one of maximization because alternative policy tools are considered to be mutually exclusive. If the maximum attained in a solution of (7) is less than zero, the industry concerned should not be protected (by any policy measure). In this context it is convenient to regard "no protection" (or "laissez faire") as a particular policy measure. There could be cases in which "no protection" is the best policy; but there could be other cases where some form of protection is optimal even when the industry gets started under "no protection." We will elaborate on the general formulation (7) below.

B. *Social Net Benefits of Infant Industry Protection Policy*

Let us now define net benefits (NB) in (7). Net benefits to society obtained from the industry under a particular policy P—$NB(P)$—comprise several components: first, changes in consumers' surplus compared to that which would prevail if there was no domestic production in the industry; secondly, profits accruing to the firm(s) in the industry; thirdly, rent accruing to the factor(s) of production;[12] fourthly, government revenues; and finally, a nonpositive component, distortion incurred outside the industry, induced by, or necessitated to finance, policy P. Now we discuss each component, in more detail. The reader is advised to review the relevant paragraphs and figures in the preceding section to substantiate the statements that follow.

The first component, changes in consumers' surplus, may be either positive, nil, or negative. In period 0, change will be negative if tariff protection is in effect, while it will be nil under a subsidy scheme. In period 1, under the assumption of no government intervention, it could be either positive or nil depending on the relative positions of the marginal cost and demand curves, on the presence or absence of sellers' competition, and on the possibility of price discrimination in a monopolistic situation. The second component, profits, could

[11] Alternatively, the absence of P_1 in (7) can be interpreted as meaning that the governmental decision in period 0 includes those measures put into effect in period 1. Here we will follow the interpretation given in the text, however, for the sake of simplicity.

[12] Note here that capital goods can be included as a factor of production. Also, if we are to treat raw materials and components explicitly, part of rent could accrue to supplier industries of those inputs. These are the celebrated backward-linkage effects. The forward-linkage effects, if any, appear as consumers' surplus.

also be either positive, nil, or negative. In period 0, they may be negative because of production and/or R & D expenses, unless the firms entering into the industry are short-term-profit maximizers. Note here that profits under the policy of "no protection" could be negative in period 0 since long-term-profit maximizers would incur temporary losses in production and/or R & D expenditures on their own. If only R & D is performed in the industry in period 0, profits will certainly be nil or negative depending on the extent to which R & D expenses are covered by subsidy. In period 1, profits will be positive. The third component, rent generated by factor demand from the industry, is either nil or positive depending on the elasticity of the supply schedule and the significance of this industry in the factor market. In the previous section we assumed that it was zero in period 0 and positive in period 1. The fourth component, government revenues, could be either positive, nil, or negative: subsidy payments are regarded as negative revenues in our definition. We follow the conventional and convenient assumption that tariff revenues are disbursed to consumers and thus regard them, dollar for dollar, as valuable as consumers' surplus. In period 1 this component is zero by definition. The last component is designed to capture the cost of a subsidy scheme which inflicts no loss of consumers' surplus in the industry to which subsidy is extended. This cost is difficult to conceptualize, not to mention measure, but cannot be ignored in the real world where an ideal, non-distorting tax scheme cannot be devised. This component is zero under no government intervention, again by definition.

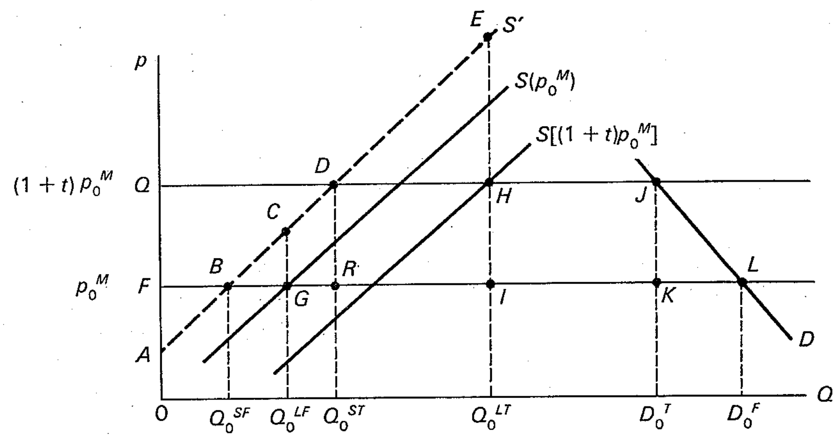
C. *The Infant Industry Argument Restated*

The formulization of the governmental decision problem in (7) can help facilitate and clarify discussions on policy issues concerning infant industry protection. The free trader's argument against protection is based on either of the following two possibilities: first, the maximum is negative; or, the positive maximum is attained under a policy of "no protection." The protectionist, on the other hand, argues that the maximum is positive and that it is realized by some policy other than "no protection." If the maximum is negative, the industry fails to pass the "Bastable test." If it is positive, but is attained under "no protectoin," then the industry fails to pass the "Kemp test." But if attaining the positive maximum requires government intervention, an authentic case for infant industry protection can be made. Note that for protection to be justified it is neither sufficient nor necessary for the domestic price to be below the supply price of imports in period 1. In any case, the "Mill test"—to the effect that "it is essential that the protection should be confined to cases in which there is ground of assurance that the industry which it fosters will after a time be able to dispense with it" [9, p. 92] —has to be passed if protection is to be justified. Below we will take up a number of cases from the previous section to illustrate the general principle discussed here.

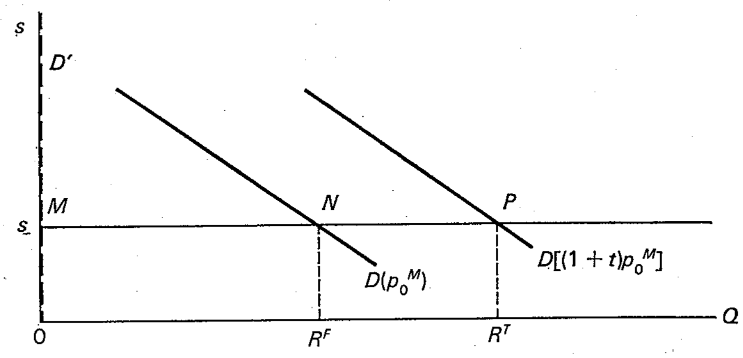
So far we have not discussed the set of policy measures from which an optimal P is selected. In principle, the set should include each and every policy measure that affects NB. Here we consider only those policy tools which change product and factor prices facing the firms, i.e., tariff, production subsidy, and R & D

Fig. 9. Effects of Tariff Protection in Period 0

a. Product market



b. R & D market



subsidy. One interesting topic not covered here is anti-monopoly regulation. Full treatment of this topic would require more detailed specification of firm behavior (including the formation of expectations about policy change) and the nature of technology (both in development and use). It seems possible, in theory at least, that the maximum in (7) could be attained by fostering a monopolistic infant, which is eventually forced to break up in its maturity to form a competitive industry.

First, we will see how a tariff or a subsidy affects firm behavior as represented by the trio of equations (4) and Figure 1 in the previous section. An ad valorem tariff of rate t raises the supply price of imports from p_0^M to $(1+t)p_0^M$. This finds its geometric expression as an upward shift of the derived-demand-for-labor curve, D', in Figure 1b. It will also shift the D curve in Figures 1b and 1c upward and the S curve in Figure 1a downward. Thus, a new equilibrium will be characterized by a set of higher values of Q^*_0, L^*_0, R^*, and therefore of L^*_1, if the firms are long-term-profit maximizers. If, on the other hand, they are short-term-profit maximizers, R & D will be nonexistent and thus—depending on R & D's importance in the "knowledge accumulation function" (2)—the level of knowledge in period 1, Z_1, may or may not be increased. If we were to accept Baldwin's claim that there are no costless learning effects [1, p. 299] ($A(Z_0, L_0, 0)=Z_0$ is one way of formalizing his idea), there would be no technological progress in the industry.

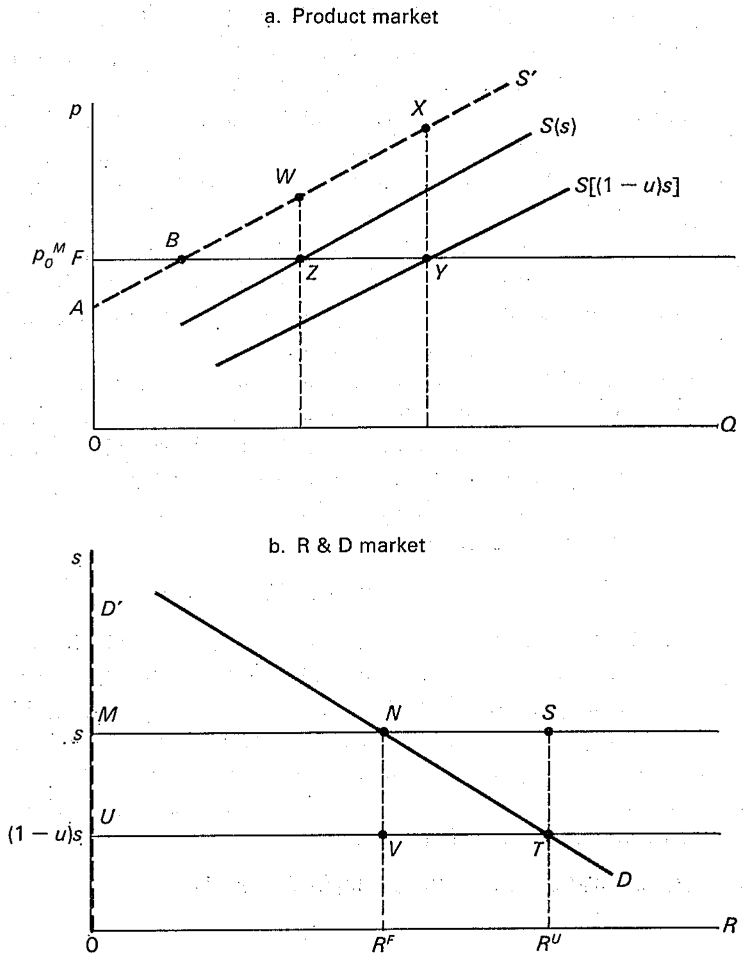
Now let us see what the components of NB_0 and NB_1 are. We reproduce Figures 1a, 1b, and 1c, adding schedules corresponding to p_0^M and $(1+t)p_0^M$ (see Figure 9). In Figure 9 and in the passages below, the superscripts S and L stand for short-term-profit maximization and long-term-profit maximization respectively, and F and T free trade and tariff protection.

Before considering the effects of tariff protection, let us take up the case of free trade, or "no protection." With no protection, equilibrium output in period 0 is Q_0^{SF} for a short-term-profit maximizing industry and Q_0^{LF} for a long-term-profit maximizing industry. The equilibrium R & D expenditures are nil for the former and R^F for the latter. The only non-zero component of NB_0 in this case is profits to the industry, Π_0, Π_0^{SF} is equal to $\triangle AFB$ and Π_0^{LF} is equal to $(\triangle AFB-\triangle BCG-\square 0MNR^F)$, which could very well be negative. The relevant components of NB_1, consumers surplus, profits, and rent, are indicated geometrically in Figures 4, 5, 7, and 8, and discussed in the paragraphs pertaining thereto. The reader is advised to review those figures and paragraphs for identification of NB_1 under various assumptions. It should be understood that under any circumstances all the components of NB_1 tend to be larger for a long-term-profit maximizing industry than for a short-term-profit maximizing industry since the former experiences a greater downward shift of the marginal cost curve (equivalent to a greater upward shift of the derived demand for labor curve).

Now let us introduce a tariff. The equilibrium outputs are now raised to Q_0^{ST} and Q_0^{LT}, while the equilibrium R & D expenditures are nil and R^T, for short-term and long-term-profit maximizing industries respectively. Let us check the value of NB_0 component by component: first, the change in consumers' surplus is negative and equal to the area $\square FQJL$; secondly, profits are $\triangle AQD$ for a short-term-profit maximizing industry and are equal to $(\triangle AQD-\triangle DEH-\square 0MPR^T)$ for a long-term-profit maximizing industry; and finally, tariff revenues to government are given by $\square RDJK$ for the former and $\square IHJK$ for the latter. Net welfare loss due to tariff protection in period 0, $(NB_0^F-NB_0^T)$, is given by $(\triangle DBR+\triangle JKL)$ for the former and $(\triangle EBI+\triangle JKL+\square R^FNPR^T)$ for the latter.

The case of a production subsidy can be traced analogously. Its effect is to raise the product price for the producers but not to the consumers. Thus the first component of NB_0—change in consumers' surplus—is nil, and the second component—profits—is altered as in the case of a tariff. Under a subsidy scheme, government payments of $\square FQDR$ or $\square FQHI$, as the case may be, have to be

Fig. 10. Effects of an R & D Subsidy in Period 0



made, and the raising of funds could cause distortions elsewhere in the economy, both factors constituting negative elements in NB_0. Thus NB_0 under a tariff will be greater than NB_0 under a production subsidy that realizes the same producers' price as the tariff, if and only if the deadweight loss in consumers' surplus is smaller than the distortion cost of raising funds to underwrite the subsidy.

We now turn our attention to the case of an R & D subsidy, expressed as a downward shift of the horizontal line in Figure 9b. The situation is depicted for a subsidy of rate u (see Figure 10). The R & D subsidy scheme does not affect a short-term-profit maximizing industry at all and therefore NB_0 remains the same (no subsidy payments are made since no firm performs R & D). To a long-term-profit maximizing industry, subsidy payments of $\square UMST$ will be made, and profits then will be $(\triangle AFB - \triangle BXY - \square OUTR^U)$, compared with the original level $(\triangle AFB - \square 0MNR^F)$. Finally, the distortion cost of raising revenues for the subsidy payments must be taken into account.

Let us now turn our attention more explicitly to the difference between private and social cost-benefit calculations; this is the central issue in the infant industry argument. In the previous section, the firm's decisional problem is formalized as (3) and the governmental decision problem as (7). Let us suppose for now that the firm's discount rate, r in (3), is equal to the social rate of discount, rs in (7), and that there are no dynamic external economies. Then the maximand in (3) constitutes one part of the maximand in (7); the other parts, other elements of NB_0 and NB_1, signify the divergence between private and social optimality. It should be noted here that the divergence defined above is determined for any given policy measure (including "no protection") and thus is conceptually distinct from the presence or absence of protection. The maximum of (3) is attained, under any given policy, through private calculations, and the rest of the maximand of (7) consists of the distortion costs of the policy in period 0 and the gains in consumers' surplus and factor rent in period 1. As has been stated previously, a government's objective is to maximize (7), but it is now clear that this maximization problem has to be solved under the constraint that the maximum of (3) is nonnegative; otherwise, what seems to be the socially optimal solution would not be realized.

D. *An Illustrative Case of Dynamic External Economies*

Here we take up the particular specification given toward the end of the previous section. Let us assume for the sake of simplicity that there are no costs of subsidy within or without the industry and no pecuniary externalities in the form of consumers' surplus or rent. Under these simplifying assumptions the only source of divergence between private (at firm level) and social (at industry level) cost-benefit calculations is dynamic externalities in the creation of knowledge.

The decision problem for firm i is as follows:

$$\max_{L_0,R,L_1} [(p_0Q_0-w_0L_0-sR)+(p_1Q_1-w_1L_1)(1+r)^{-1}], \qquad (8)$$

subject to

$$Q_t=Z_tf(L_t), \quad t=0,\ 1,$$

and

$$Z_1^i=\sum_{j=1}^{N} \lambda_{ji}A^j(Z_0^j,\ L_0^j,\ R^j).$$

The decision problem for government is to maximize the sum of the maximand in (8) (summed over N firms). Under the symmetry assumption expounded in relation to equation (6) in the previous section, socially optimal levels of L_0, R, and L_1 (the same for all firms) are determined as the solution of the following decision problem:

$$\max_{L_0,R,L_1} N[(p_0Q_0-w_0L_0-sR)+(p_1Q_1-w_1L_1)(1+r)^{-1}], \qquad (9)$$

subject to

$$Q_t=Z_tf(L_t), \quad t=0,\ 1,$$

and

$$Z_1=[\lambda+(N-1)\mu]A(Z_0,\ L_0,\ R).$$

The first-order conditions are as follows:

$$[p_0Z_0f'(L_0)-w_0]+[\lambda+(N-1)\mu]p_1f(L_1)\frac{\partial A}{\partial L_0}(1+r)^{-1}=0, \tag{10a}$$

$$-s+[\lambda+(N-1)\mu]p_1f(L_1)\frac{\partial A}{\partial R}(1+r)^{-1}=0, \tag{10b}$$

$$[p_1Z_1f'(L_1)-w_1](1+r)^{-1}=0. \tag{10c}$$

This social optimum will be attained if firms foresee the give and take in knowledge creation correctly. Suppose, however, that firms are pessimistic and act under the worst-case assumption that they give but do not take. Then the (subjective) "knowledge accumulation function" for firm i is:

$$Z_1=A(Z_0,\ L,\ R), \tag{11}$$

which is a special case of the function in (8). The first-order conditions under this assumption are obtained by replacing $[\lambda+(N-1)\mu]$ in the trio of equation (10) by λ. The optimal private levels of L_0, R, and L_1 will then tend to be smaller than what they would be under the socially optimal resource allocation. Let us illustrate this for the R & D market:

$$\text{Private solution:}\quad p_1\lambda(1+r)^{-1}f(L_1)\frac{\partial A}{\partial R}=s, \tag{12a}$$

$$\text{Socially optimal solution:}\quad p_1\lambda(1+r)^{-1}f(L_1)\frac{\partial A}{\partial R}=\frac{\lambda}{[\lambda+(N-1)\mu]}s. \tag{12b}$$

Note here that the optimal level of L_1 is a function of R (as well as a function of L_0, which is here assumed to be fixed at the optimal level). To attain the socially optimal level of R, a subsidy of rate $u=(N-1)\mu/[\lambda+(N-1)\mu]$ will be required, i.e., $(1-u)s=\{\lambda/[\lambda+(N-1)\mu]\}s$. As was indicated above, the presence of dynamic external economies could justify government intervention. The final conclusion must be reached by weighting the increase in social welfare against the increase in distortion costs brought about by a protective policy.

V. CONCLUSION

In this paper we have reformulated the infant industry argument as a decisional problem (a constrained maximization problem) for government (the general formula is shown as (7) in Section IV). Theories of firm behavior and market equilibrium have been explicitly expounded. Many of the assumptions adopted (especially those concerning the formation of expectations) are problematic, it is true; we believe, however, that we have succeeded in clarifying the nature of issues raised by Kemp [7], Grubel [4], Negishi [10], Baldwin [1], Johnson [6], and others. The social cost-benefit analysis framework, formally applied to this field by Grubel [4] and also utilized in the appendix to Negishi [10], has proved to be very useful in the formulation of the maximization problem for government.

It is time to evaluate the usefulness of the theoretical approach we have constructed.

Herbert G. Grubel, after giving an explicit cost-benefit analysis expression to the "Bastable test," offered the following negative conclusion with regard to the theoretical approach he initiated: "With the tools of economic analysis presently available it appears to be impossible to identify industries qualifying for infant protection both ex ante and ex post" [4, pp. 338–39]. The lapse of almost fifteen years since Grubel wrote these words does not seem to have improved the situation of economic analysis drastically enough to make him change his mind, despite the recent flourishing of social cost-benefit analysis both in theoretical articulation and in application. The nature of the basic difficulty in the choice of industry to be protected is concisely expressed by J. E. Meade: "A decision about a structural change of policy is bound to remain a much more hit-or-miss affair than a decision about a marginal change—which itself will always be a matter of great uncertainty" [8, p. 134]. Meade's dictum remains as true today as it was at the time of his writing, although the standard procedures for the treatment of risk and uncertainty have been considerably developed by cost-benefit analysts. Thus, the governmental problem might better be formulated in a statistical decision theory framework taking the loss function explicitly into consideration. Grubel's proposal for a self-terminating scheme touched on in the previous section should be evaluated properly in this light. One way will be to put to reformulate the infant industry argument along this line in the framework of a sequential decision-making model.[13] For now, however, let us conclude our paper with another passage from J. E. Meade:

> In practice, of course, it will be impossible to make any accurate calculation on these lines; but it is only by consideration of the factors involved in the calculations that an informed "hunch" can be reached as to the desirability of setting up the new industry. [8, p. 255]

[13] Placed in the context of statistical decision theory, the infant industry argument will be formulated as a methodology for determining weights to be attached to two types of errors, i.e., the error of protecting an unqualified industry and the error of not protecting a qualified industry.

REFERENCES

1. BALDWIN, R. E. "The Case against Infant-Industry Tariff Protection," *Journal of Political Economy*, Vol. 77, No. 3 (May–June 1969).
2. BECKER, G. *Human Capital* (New York: National Bureau of Economic Research, 1964).
3. CORDEN, W. M. *Trade Policy and Economic Welfare* (London: Oxford University Press, 1974).
4. GRUBEL, H. G. "The Anatomy of Classical and Modern Infant Industry Arguments," *Weltwirtschaftliches Archiv*, Vol. 97, No. 4 (December 1967).
5. JOHNSON, H. G. "Optimal Trade Intervention in the Presence of Domestic Distortions," in *Trade, Growth and the Balance of Payments*,, ed. R. E. Caves, P. B. Kenen, and

H. G. Johnson (Chicago: Rand McNally; Amsterdam: North-Holland, 1965).
6. ———. "A New View of the Infant Industry Argument," in *Studies in International Economics*, ed. I. A. McDougall and R. Snape (Amsterdam: North-Holland Publishing Co., 1970).
7. KEMP, M. C. "The Mill-Bastable Infant-Industry Dogma," *Journal of Political Economy*, Vol. 68, No. 1 (February 1960).
8. MEADE, J. E. *Trade and Welfare* (London: Oxford University Press, 1955).
9. MILL, J. S. *The Principles of Political Economy*, ed. Ashley (London: Longmans, Green and Company, 1909).
10. NEGISHI, T. "Protection of the Infant Industry and Dynamic Internal Economies," *Economic Record*, Vol. 44, No. 105 (March 1968).
11. NORDHAUS, W. D. *Invention, Growth and Welfare* (Cambridge, Mass.: MIT Press, 1968).
12. PENROSE, E. T. *The Theory of Growth of the Firm* (Oxford: Basil Blackwell, 1959).
13. ROSEN, S. "Learning by Experience as Joint Production," *Quarterly Journal of Economics*, Vol. 86, No. 3 (August 1972).

APPENDIX

In this appendix we denote quantities relating to the industry by upper-case letters and those relating to the firm by lower-case letters. Prices continue to be denoted by lower-case letters

The following schedules determine the product price, p, and the factor price, w, simultaneously:

Labor supply:	$L^S=S(w)$	Labor market
Labor demand:	$L^D=E(w, p)$	
Product supply:	$Q^S=F(p, w)$	Product market
Product demand:	$Q^D=D(p)$	

Equilibrium is reached when $L^S=L^D$ and $Q^S=Q^D$. The relationship between p and w that keeps the labor market in equilibrium is implicitly defined by the equation, $S(w)=E(w, p)$ and the counterpart for the product market by the equation, $D(p)=F(p, w)$. The slopes of these curves are:

$$\left.\frac{dw}{dp}\right|_L=\frac{E_p}{S_w-E_w}$$

and

$$\left.\frac{dw}{dp}\right|_Q=\frac{D_p-E_p}{F_w},$$

where subscripts L and Q denote labor market and product market respectively, and E, S, and D, with subscripts, signify the partial derivative of those functions with respect to the variable represented by the subscript.

Since

$$E(w, p)=\sum 1\left(\frac{w}{p}\right) \quad \text{and} \quad F(p, w)=\sum q\left(\frac{p}{w}\right);$$

$$E_w=\sum 1_w=\sum 1'\cdot\frac{1}{p}=\frac{1}{p}\sum 1',$$

$$E_p=\sum 1_p=\sum 1'\cdot\left(-\frac{w}{p^2}\right)=\left(-\frac{w}{p^2}\right)\sum 1',$$

$$F_p=\sum q_p=\sum q'\cdot\frac{1}{w}=\frac{1}{w}\sum q',$$

and

$$F_w=\sum q_w=\sum q'\cdot\left(-\frac{p}{w^2}\right)=\left(-\frac{p}{w^2}\right)\sum q'.$$

Therefore,

$$\max\frac{dw}{dp}\bigg|_L=-\frac{E_p}{E_w}=\frac{w}{p}=-\frac{F_p}{F_w}=\min\frac{dw}{dp}\bigg|_Q.$$

This proves that, around equilibrium, the slope of the curve for labor market equilibrium is less steep than that of the curve for product market equilibrium in Figure 3. Note the two extreme cases; $S_w=\infty$ makes the former horizontal and $D_p=\infty$ makes the latter perpendicular.

Shifts in the labor supply function, $S(w)$, or in the product demand function, $D(p)$, result in different pairs of p^* and w^* as equilibrium prices. The equilibrium industry supply schedule and equilibrium industry derived demand schedule are derived in Figure 4 as loci that satisfy the relationship between the equilibrium prices.