AN ECONOMIC ANALYSIS OF THE DEVELOPMENT OF
SUBSTITUTES WITH SOME ILLUSTRATIVE EXAMPLES
AND IMPLICATIONS FOR THE BEEF INDUSTRY

by

John Spriggs
AN ECONOMIC ANALYSIS OF THE DEVELOPMENT OF SUBSTITUTES
WITH SOME ILLUSTRATIVE EXAMPLES AND
IMPLICATIONS FOR THE BEEF INDUSTRY

by

John Spriggs*

The development of substitutes poses an important problem to agricultural producers throughout the world. However, the main sufferers are probably those in underdeveloped countries for which the natural products provide a major source of exchange earnings. Take, for example, the case of rubber. Over 75 percent of the world supply originates from less developed countries in Southeast Asia. On the other hand, synthetic rubbers are produced mostly in developed countries such as the U.S.A., Canada and Germany. There are now synthetic rubbers which apparently can duplicate or supercede all the qualities of natural rubber. In important end-uses such as car tires, synthetic rubber completely dominates the market.

What is a substitute and what do we mean by development? In this paper a substitute is defined as a product that provides a combination of qualities similar to those found in an agricultural product but is manufactured out of different, often non-agricultural raw materials. This includes such products as margarine and rayon which have an agricultural

*Research assistant, Department of Agricultural and Applied Economics, University of Minnesota. The author wishes to acknowledge the very substantial help of Professor James Houck throughout this study, but claims full responsibility for any errors that may remain.
base as well as those such as nylon with a non-agricultural base.

Development is another term that requires definition. A common notion is to distinguish "research and development." However, in this paper such a distinction is not made. Development is used here to describe the dynamic process by which improvement occurs, either in the new product's marketable qualities or in the techniques used in its production.

There are three main parts to this paper. The first concerns the adaptation of conventional demand and supply theory to explain why and when the development of substitutes occurs. The second part is a discussion of four examples of substitutes that appear to have completed their development process. This is to see how the theory fares in practice. In the third part, implications of the theory are drawn for the beef industry which recently had its first serious encounter with substitutes.

I. THEORY

Theories on product development are not new. Schumpeter [19], a pioneer in this field, emphasized that development (in a capitalistic system) is propelled by the force of invention. He considered inventions to be an exogenous force which would "shock" the economic system into an upswing. Schmookler [17] however, found that inventions rather than leading an upswing were in fact caused by it. By analyzing the number of patents issued from year to year he found that they rose and fell in lagged response to upswings and downswings in the economy.

Many empirical economists have attempted to incorporate the ideas of either Schumpeter or Schmookler into conventional economic analysis.
These attempts have centered on the use of a production function framework. Thus, development is represented by shifts in a production function.

However, in the present problem at least the production function approach seems unsuitable. The production function may be useful in describing the development process but conceals the motivating forces for development. In particular, the development of substitutes is postulated to depend largely on conditions existing in the market for the natural product. However, merely examining shifts in the production function for a substitute tells us nothing of this relationship. In this paper then, the production function framework is not used. Rather the problem of the development of substitutes is cast into a price analysis framework.

Suppose a substitute's development is found to be technologically feasible. Further research beyond the exploratory stage will only be undertaken if the expected benefits outweigh the expected costs. And because a high risk element is usually involved the expected benefits should far outweigh the expected costs. Schmookler [17] said that every invention is a fixed cost but that the benefits to be derived from it vary with socio-economic changes. However, when the invention is also a substitute, the expected benefits also vary with the price, quality and market size of the natural product. For a natural product with a

\[1\] Such changes as he listed were urbanization, declining family size, the changing status of women, changes in relative factor costs and increases in population and per capita income.
given market size and quality, the higher its price, the higher the expected benefits from developing the substitute.

Thus, the development of substitutes is not a smooth or continuous process over time. Some of the "bumpiness" is due to the uneven acquisition and application of new technical knowledge. However, much of it is attributable to conditions existing in the market for the natural product. High prices for the natural product are important in the timing of development. They are an inspiration to entrepreneurs and researchers alike. In addition, they increase consumer awareness about the product and make him more receptive to a cheaper or improved quality substitute.

In the analysis to follow we shall examine the development of substitutes first in a static framework, then in a dynamic framework.

**A Static Approach to Development**

In the first and simplest of models consider a natural product A and a potential substitute B. We may assume that a certain body of knowledge exists concerning B, so that its potential developers have a "perceived" supply function for that good, following development. Let us assume for now that A and B are of equal quality in the eyes of the potential developer. Then he also perceives that the initial demand curve for A is equivalent to the demand curve for A plus B. We postulate that the supply curve for A lies below that for B. By this we are assuming initially that any given quantity of the natural product is available at a lower price than the substitute. Then we may have the situation presented in figure 1.
In this figure, \( D(A + B) \) is the "perceived" demand curve for the combination of qualities offered by both the natural product and its substitute. \( S(A) \) and \( S(B) \) are the supply curves respectively for the natural product and its substitute. \( S(A + B) \) is the horizontal summation of \( S(A) \) and \( S(B) \) and represents total supply at any given price.

It should be apparent that in this situation no development of \( B \) will occur since the demand can be fully satisfied by \( A \) at a lower price than would need to be charged by the producers of \( B \) to cover costs.

Let us consider three basic cases in which the development of \( B \) may occur. [Note: In this exposition we shall omit the term "perceived" to describe the demand and supply of \( B \). This is not to lessen the
importance of the notion but is to avoid repetition of a point which has already been made.]

1. An increase in demand for $A$ plus $B$—This may result from changes in the usual determinants of demand: population size, consumer income, prices and availability of other goods and services, and consumer tastes and preferences. (Some of these correspond to Schmookler's socio-economic changes; see footnote 1.) Given the situation as in figure 1, suppose consumer income is increased as a result of expansionary monetary and fiscal policy. Then provided $A$ and $B$ are "normal" goods, we would expect an outward shift in the demand curve $D(A + B)$. The equilibrium price and quantity will both increase. Above a certain price it will be economically worthwhile for $B$ to be produced. Thus, in figure 2(a) if the demand curve shifts out from $D(A_1 + B_1)$ to $D(A_2 + B_2)$, $Q_1Q_T$ of $B$ and $OQ_1$ of $A$ will be produced. Since $A$ and $B$ are of equal quality the equilibrium price will be the same ($P^*$) but higher than initially ($P_0$).

![Figure 2(a)](image-url)
2. A decrease in the supply of A—This may result from increases in input prices, the profitability of products that compete for the same resources as A, institutional constraints, and bad weather and pest conditions. Suppose a change in one or more of these shifts the supply curve for A from $S(A_1)$ to $S(A_2)$ as in figure 2(b). Then the total supply curve $S(A_1 + B_1)$ shifts to $S(A_2 + B_1)$. Now the equilibrium price increases sufficiently to call forth the development of B. At the new price $P^*$, $Q_2Q_T$ of B and $QQ_2$ of A are produced. As in the first case the equilibrium price is higher than in the initial situation. Now, however, the effect on the quantity of A demanded is different. In the first case more A was demanded than initially, while in the present case the demand for A is less than initially.

![FIGURE 2(b)](image)
3. A downward shift in the supply curve for B—This may result from the same factors as caused a decrease in the supply of A. But of course now they are operating in the opposite direction. In addition, an exogenous technological advance in B may cause its supply curve $S(B)$ to shift down.

We can either think of $S(B)$ as shifting down or of the total supply curve $S(A + B)$ sliding down $S(A)$. In figure 2(c) we have depicted the total supply curve sliding down $S(A)$ from $S(A + B_1)$ to $S(A + B_2)$. This new kinked supply curve intersects with the total demand curve on its flatter segment. This implies that producers of B are able to cover average and marginal costs at the prevailing market price ($P_0$), and that it will be profitable for them to develop B. As they commence production the equilibrium price will decline to $P^*$. B's share of the market $Q_3Q_T$ derives partly from the expansion in demand as a result of the price fall but also it derives in part from the displacement of A from $OQ_0$ to $OQ_3$. Thus, at the new equilibrium position less A will be demanded than initially.

![Figure 2(c)](image-url)
Endogenous Technological Advance

Once production of B is initiated, further development may be induced by endogenous technological advance. This is particularly important where the initiating force is only temporary, such as a crop failure (case 2). It may come about in either of the following ways.

With the production of B, new possibilities for cost reduction may arise which previously were unforeseen. These come as a result of the process "learning by doing." In addition, if the initial production had proved successful in penetrating the market for A the entrepreneur may be induced to bring about further cost reductions through a larger scale operation. Both "learning by doing" and "economies of scale" result in shifts down (and to the right in the latter case) of the supply curve for B.

For example, suppose a crop failure shifts the total supply curve from $S(A_1 + B_1)$ to $S(A_2 + B_1)$ as in figure 3. This is sufficient to initiate production of B. Subsequently, suppose endogenous technological advance causes the supply function for B to shift down. Then even when A recovers from its temporary supply cutback (supply curve for A returns to $S(A_1)$) the production of B may continue. The endogenous technological advance causes the total supply curve $S(A_1 + B_1)$ to slide down $S(A_1)$. If as in figure 3 it slides down to $S(A_1 + B_2)$ so as to intersect with D(A + B) on its (the supply curve's) flatter segment then production of B will continue. And, what is more, further development may be induced.
The development process may involve increases in the quality of B. Thus far we have assumed A and B to be of equal quality in the eyes of the entrepreneur (and the consumer). In reality, this is rarely the case. Usually a substitute is initially inferior to the natural product, if not physically at least in the mind of the consumer. However, development of the substitute frequently takes on the form of quality improvement as well as cost reductions.

To incorporate quality changes into our model, let us first consider a simple algebraic and graphical model in which the focus is on the cross-elasticities of demand between A and B. The model is:

(1) \( Q_{DA} = a_0 - a_1 P_A + a_2 P_B \) \hspace{1cm} (demand for A)

(2) \( Q_{SA} = b_0 + b_1 P_A \) \hspace{1cm} (supply of A)

(3) \( Q_{DB} = c_0 + c_1 P_A - c_2 P_B \) \hspace{1cm} (demand for B)
$$(4) \quad QSB = d_0 + d_1 PB \quad \text{(supply of } B)$$
$$\quad (5) \quad QDA = QSA \quad \text{(market clearing identity for } A)$$
$$\quad (6) \quad QDB = QSB \quad \text{(market clearing identity for } B)$$

where $PA$ and $PB$ are prices of $A$ and $B$ respectively, and $QDA$, $QSA$, $QDB$ and $QSB$ are the quantities demanded and supplied of $A$ and $B$ respectively.

Let us assume all variables are in logarithms so that the coefficients are interpreted as direct- and cross-elasticities. From this six equation model the equilibrium prices and quantities of $A$ and $B$ may be determined and expressed in terms of the parameters only. They are:

$$\quad (7) \quad PA_{Eq} = \frac{(a_0 - b_0)(c_2 + d_1) + a_2 (c_0 - d_0)}{(a_1 + b_1)(c_2 + d_1) - a_2 c_1}$$
$$\quad (8) \quad QA_{Eq} = b_0 + b_1 \left[ \frac{(a_0 - b_0)(c_2 + d_1) + a_2 (c_0 - d_0)}{(a_1 + b_1)(c_2 + d_1) - a_2 c_1} \right]$$
$$\quad (9) \quad PB_{Eq} = \frac{(a_1 + b_1)(c_0 - d_0) + c_1(a_0 - b_0)}{(a_1 + b_1)(c_2 + d_1) - a_2 c_1}$$
$$\quad (10) \quad QB_{Eq} = d_0 + d_1 \left[ \frac{(a_1 + b_1)(c_0 - d_0) + c_1(a_0 - b_0)}{(a_1 + b_1)(c_2 + d_1) - a_2 c_1} \right]$$

We are interested in the effects of an improvement in the quality of $B$ relative to $A$ on our model. A priori we would expect $c_1$ (the cross-elasticity of demand for $B$ with respect to a change in the price of $A$) to increase. This means, if the quality of $B$ improves, the quantity of $B$ demanded will increase more rapidly with a given increase in $PA$. Using the approximate Slutsky cross-elasticity symmetry relation, we also expect $a_2$ to increase. This would reinforce the effects of an increase in $t_1$, on the dependent variables of equations (7) to (10).\(^2\) Taking $\quad 2/$Using the exact Slutsky relation it is conceivable that $a_2$ would
the total differential of equation (7) where \( a_2 \) and \( c_1 \) are assumed to increase we have:

\[
dPA_{\text{Eq}} = \frac{a_2 (NM + a_2 K)}{(LM - a_2 c_1)^2} \quad dc_1 + \frac{M (KL + c_1 N)}{(LM - a_2 c_1)^2} \quad da_2
\]

where \( K = c_0 - d_0 \)
\( L = a_1 + b_1 \)
\( M = c_2 + d_1 \)
\( N = a_0 - b_0 \)

We wish to know whether \( dPA_{\text{Eq}} \) is positive or negative. Since \( a_2, M \) and \( (LM - a_2 c_1)^2 \) are all positive the problem is to determine the signs of \( (NM + a_2 K) \) and \( (KL + c_1 N) \). Because of the minus signs in \( K \) and \( N \) it is conceivable that one or both expressions are negative. However, as we shall see presently in the graphical analysis, a negative sign on these expressions cannot occur if we ignore income effects. Hence, ignoring income effects, \( dPA_{\text{Eq}} \) is positive, and we expect that an improvement in the quality of B will increase the equilibrium price of A.

Let us turn now to the effect of such an improvement on \( QA_{\text{Eq}} \). From equation (8)

\[
dQA_{\text{Eq}} = b_1 \ dPA_{\text{Eq}}
\]

decrease. From the Slutsky relation, \( da_2 = r_B \left[ \frac{dc_1}{r_A} + d\eta_B - d\eta_A \right] \) where \( r_A \) and \( r_B \) are the expenditure proportions and \( \eta_A \) and \( \eta_B \) are the income elasticities of A and B respectively. Thus, \( da_2 \) will be negative only if the improved quality of B has a sufficiently large impact on the income elasticity of A.
Thus if \( dP_{AEq} \) is positive so also is \( dQ_{AEq} \). The improvement in B leads to a higher equilibrium quantity in the market for A.

The effect on \( PB_{Eq} \) can be determined by taking the total differential of equation (9).

\[
dP_{BEq} = \frac{L(NM + a_2K)}{(LM - a_2c_1)^2} \cdot dc_1 + \frac{c_1(KL + c_1N)}{(LM - a_2c_1)^2} \cdot da_2
\]

As in the earlier case, zero-income effects imply that \( NM + a_2K > 0 \) and \( KL + c_1N > 0 \). Hence \( dP_{BEq} \) is positive. Since \( dQ_{BEq} = d_1 \cdot dP_{BEq} \) it follows that \( dQ_{BEq} \) is also positive.

The effects of quality improvements in B may also be explained graphically. Using equations (1) to (6) we can derive

\[
(11) \quad P_{AEq} = \frac{a_0 - b_0}{a_1 + b_1} + \frac{a_2}{a_1 + b_1} \cdot PB
\]

which is the equation of the locus of equilibrium points in the market for A given any price of B; and

\[
(12) \quad P_{BEq} = \frac{c_0 - d_0}{c_2 + d_1} + \frac{c_1}{c_2 + d_1} \cdot PA
\]

which is the equation of the locus of equilibrium points in the market for B given any price of A. Equations (11) and (12) may be represented graphically on the PA-PB plane as in figure 4. In order to obtain a meaningful simultaneous equilibrium in both markets we require that the curves \( AA_1 \) and \( BB_1 \) (representing equations (11) and (12) respectively) intersect in the positive quadrant. That intersection point yields the two equilibrium prices, from which the equilibrium quantities are simply determined. Hicks [10, p. 69] has shown that for stability \( AA_1 \) must
intersect $BB_1$ from below. In addition he shows that if income effects may be ignored then it is not possible to have an unstable system.

Using the result: slope $AA_1 >$ slope $BB_1$, we have

$$\frac{a_1 + b_1}{a_2} > \frac{c_1}{c_2 + d_1}.$$  Thus, $(a_1 + b_1)(c_2 + d_1) - a_2c_1 > 0$ or $LM - a_2c_1 > 0$.

Now, in order to achieve equilibrium within the positive quadrant we also require that the intercept for $BB_1$ exceed that for $AA_1$. Thus,

$$\frac{c_0 - d_0}{c_2 + d_1} > \frac{-(a_0 - b_0)}{a_2}.$$  Rewriting, this becomes $a_2(c_0 - d_0) + (a_0 - b_0).

$(c_2 + d_1) > 0$ or $(NM + a_2K) > 0$. This expression is one of the two whose sign was ambiguous. The sign of the other expression $(KL + c_1N)$ may be determined from the two inequalities we have just established: $LM - a_2c_1 > 0$ and $NM + a_2K > 0$. Since $LM - a_2c_1 > 0$, then $a_2 < LM/c_1$.

Further, since $NM + a_2K > 0$, then clearly $NM + \frac{LM}{c_1}K > 0$. Therefore, $c_1NM + LMK > 0$, and as $M$ is positive, $c_1N + LK > 0$.

We have shown that the expressions $(LM - a_2c_1)$ and $(c_1N + LK)$ both have positive sign. Hence, as stated earlier, an improvement in the quality of $B$ will lead to an increase in the equilibrium prices of both $A$ and $B$, if we ignore income effects.
It may be instructive to further explore the graphical analysis. For simplicity let us assume for now that as the quality of B improves, $c_1$ increases but $a_2$ remains constant. In figures 4 and 5, the curves $AA_1$ and $BB_1$ are drawn so that they intersect, at point $a$. Points to the right of $AA_1$ lie in the area of excess supply of A. Such points indicate that $PA_{Eq}$ is too high given $PB$. Contrariwise, points to the left of $AA_1$ lie in the area of excess demand in the market for A. Analogously, since $BB_1 = PB_{Eq} | PA$, points above $BB_1$ indicate that B is in excess supply, while points below $BB_1$ indicate excess demand. The directional arrows in figure 4 show which way the system tends from any given point. Suppose that equilibrium in the two markets is disturbed from point $a$. The price of A rises momentarily and we arrive at b (see figure 5). At this point, A is in excess supply while B is in excess
demand. There is a tendency for PB to rise, which enhances the likelihood that development of B will occur. Suppose development does occur and it results in a quality improvement of B. This causes $c_1$ to increase so that the slope of $BB \left( \frac{c_1}{c_2 + d_1} \right)$ in equation (12) also increases.

Suppose the new locus of equilibrium points in the market for B is $BB_2$. PB increases until we reach a new equilibrium price for B at point c. Here though A is still in excess supply and we proceed in a stepwise fashion toward a new simultaneous equilibrium point d. Note that at d, the equilibrium prices for both A and B are higher than at the initial equilibrium a.
A Dynamic Approach to Development

Thus far we have used static market equilibrium models to indicate why the development process of a substitute begins. We also used a static model to examine the impact of endogenous technological change on the development process. Since development is essentially a dynamic process we may gain more insight into the process by constructing a dynamic model of the behavior of a single entrepreneur. To do this we must first go back to the entrepreneur's "perceived" demand and supply curves for the substitute, B. Suppose we are at the stage where a body of knowledge exists regarding B but that B is not yet a marketable commodity. The market for A is, let us say, as described in figure 6a.

FIGURE 6
An entrepreneur interested in developing B may perceive the demand curve for B as follows. He figures that the quality differences between A and B (both real and imagined) will mean that the maximum price he could charge consumers would be \( (P_{A_{eq}} - \Delta) \). Further, for output levels that are small relative to the total market he thinks he can expand output without lowering price. Thus in figure 6(b) the entrepreneur perceives a horizontal demand curve over some output levels 0 to \( Q_2 \).

Given the body of knowledge concerning B the entrepreneur also perceives an average cost curve for B or at least the minimum point on this curve. Note that this curve may or may not turn out to be the actual average cost curve. The important thing at this stage is what the entrepreneur perceives not what actually happens.

Using the familiar equilibrium theory of a firm in a competitive market we presume that the entrepreneur will develop the substitute providing the perceived minimum average cost is no greater than the perceived horizontal demand curve. Since the position of B's demand curve depends on the equilibrium price of A, we may specify a "critical price" for A, where, critical price (\( PC \)) = minimum average cost of B + \( \Delta \). This is illustrated in figure 7.
Development is presumed to occur when the expected price of A exceeds this critical price. The higher the price of A above this critical price the greater the potential "economic profit" from investing in B. In other words, investing in B becomes relatively more attractive than investing in the competing alternatives. And the entrepreneur will tend to reassess his investment portfolio, giving more weight to B.

Postulate 1: The higher the price of A above its critical price, the greater the investment in B.

With successful initial development the critical price may be expected to decline as a result of:

(a) Endogenous technological change. This could not be perceived by the entrepreneur prior to the initial development. With reference to figure 7 this factor may cause the A margin to decline (quality improvements in B) or may cause the perceived average cost curve to shift down (lower production costs for B).

(b) Declining risk. With development the entrepreneur has greater knowledge about B and its impact on the market. This tends to lower the perceived average cost curve in figure 7.

There are other factors which affect the critical price. They include the perception errors of the entrepreneur: the perceived price discount (Δ) or the perceived average cost curve may turn out to be too high or too low. These, unlike the first-mentioned factors, may go in either direction. Ignoring such mistaken perceptions let us postulate the following.

Postulate 2: The greater the size of investment in B during a given period, the greater the decline in critical price.

Here we first imply that the larger the investment the greater the endogenous technological change. This is consistent, though on the micro
level, with Schmookler's hypothesis that inventions rise and fall in lagged response to upswings and downswings in the economy's level of investment expenditure. We also imply that the greater the investment the greater the decline in risk. This seems reasonable: the more you invest in something, the more knowledge you expect to have of its potential in the market.

Applying postulates 1 and 2 we obtain the result:

The higher the price of A above the critical price, the greater the reduction in the critical price. Hence, other things equal, the greater also will be the tendency for subsequent development.

This result is important for the following analysis. The implied relationship is shown in figure 8. [Although the linear functional form is used this is not meant to imply that the real world is linear. This form was chosen merely for simplicity of exposition.] Two further simplifying assumptions apply throughout this analysis: (1) If development is possible and desired by the entrepreneur, it is assumed to occur without lag; and (2) Any effect on the critical price is assumed to occur in the period following a price change in A. Let us call the relationship depicted in figure 8 a "response function." Algebraically this function may be stated:

\[ \Delta PC(t) = b_t [PA(t) - PC(t)] \]

where

- \( PC(t) = \) critical price in period \( t \)
- \( PA(t) = \) price of A in period \( t \)
- \( \Delta PC(t) = \) change in critical price = \( PC(t) - PC(t + 1) \)
If we begin with \( PA = PA(1), \) \( PC = PC(1) \) and the response coefficient \( (b_t) \) we may determine the change in critical price. Note that if we begin with \( PA = PA(1)^\ast \) which is below the critical price then there is zero change in the critical price. This is simply because at this price of \( A \) there is no development.

\[
\Delta PC(t) = b_t [PA(t) - PC(t)]
\]

**FIGURE 8**

To explain the response over more than one time period this figure becomes cumbersome. We have to take the amount off the vertical axis, \( \Delta PC(t) \) and subtract this from the intercept on the horizontal axis, \( PC(t) \)
to obtain a new intercept for the next period. To avoid this, consider
the following approach.

The response function is:

$$PC(t) - PC(t + 1) = b_t [PA(t) - PC(t)]$$

$$\therefore PC(t) = \frac{1}{1 + b_t} PC(t + 1) + \frac{b_t}{1 + b_t} PA(t)$$

In PC-PA space this is a family of lines with slope $\frac{b_t}{1 + b_t}$ and
intercept $\frac{1}{1 + b_t} PC(t + 1)$ which varies with changes in $PC(t + 1)$.

In figure 9, this response function is relevant for any point in
the half-space to the right of the 45° line, the locus of points satis-
fying $PC(t) = PA(t)$. In this half-space $PA(t) > PC(t)$ so that development
if possible is desired and presumed to take place. Any point to the left
of the 45° line is one which leads to no development. In terms of figure 8
it is a point such as $PA(1)$. Hence, the appropriate response function
for points in this set is $PC(t) - PC(t + 1) = 0$. Let us take any point $M$
to the right of the 45° line. The entrepreneur may find himself at this
point as a result of the forces discussed earlier in the static models.
These include appropriate shifts in the supply and demand curves for A
and exogenous technological change in B. At $M$ the entrepreneur is induced
to develop B. In doing so, the critical price $[PC(0)]$ declines at a
rate determined by $b_t$. The new critical price $[PC(1)]$ is given by the
point of intersection of the response function with the 45° line. This
is shown in the appendix.

If we assume that both $b_t$ and the price of A remain constant from
year to year then in year 2 the critical price declines to $PC(2)$. The
FIGURE 9

PC(t) = PA(t)

\( b_t/(1 + b_e) \)

\( 45^\circ \)
entrepreneur has a greater incentive to develop now than he did in the first period. This situation is clearly unstable, since the decline in the critical price becomes progressively larger. We should note that the assumption about the price of A is quite plausible in the real world. Past examples of this have been government-sponsored guaranteed price schemes. What is to prevent the critical price from declining to zero? There are two possible answers.

First, the output of B may become large relative to the size of the total market. Thus, in terms of figure 6(b), output extends beyond 0Q₂. Then the entrepreneur perceives that he can only expand the output and sale of B further by lowering its price. One way of viewing this is to say the critical price must increase given the average cost curve for B. This possible answer opens up a whole new area of inquiry into the development process, but will not be further discussed in the present paper.

Instead let us turn to the second possible answer.

The factors causing the critical price to fall will probably decline in importance as the development process proceeds. Let us examine these in turn.

(a) We might expect that most of the contributions from endogenous technological advance will occur early. This is suggested because researchers, like many people, have a strong sense of discovery. Researchers, given say five years to work with a newly developed product, will learn almost all that there is to know in perhaps the first two years. In the subsequent three years there is little left to discover and hence, little contribution that can be made to lowering the critical price. Of course, there is the possibility of a "technological breakthrough" at
any time. This opens up a whole new area of discovery and the process starts all over again.

(b) We would expect the effect of **declining risk** on the critical price to itself decline. There is some lower bound beyond the risk cannot further fall. Thus, the effect on critical price either declines asymptotically to zero over time, or reaches zero when the lower bound is met.

(c) Any effect on the critical price via mistaken perception is likely to decline over time as more knowledge about the substitute is obtained.

In terms of our response function the foregoing implies that $b_t$ is large at the start of the development process, but declines toward zero at the end of the process. One possible shape for the $b$ function is shown in figure 10(a). A technological breakthrough during the development process shifts the $b$ function up, as exemplified in figure 10(b). Here the breakthrough occurs in period 1, and its effect on $b_t$ is felt in the second and subsequent periods.

![FIGURE 10]
As an illustration of how we combine the type of information in figure 10 with that in figure 9, consider the following. Suppose in period 0 that the entrepreneur perceives a critical price for A of $10. Further suppose that the actual price in that period rises above the critical price to $12. This is sufficient to initiate development. Once production commences let us assume that further development occurs at the rate described in figure 10(a). Figure 11 describes the total development process. At the end of the process the critical price is $7.8.

For a subsequent development process to commence, we would require that \( b_t \) again become positive. This can occur only through a technological breakthrough. Given such a breakthrough, the critical price immediately decreases by a certain amount, so that if the resulting point is below the 45° line, development will commence, incorporating this new knowledge. In addition, subsequent development (e.g., via endogenous technological change) will be encouraged at a rate given by the new \( b \) function.

The example in figure 11 illustrates what I have been calling a "development process." This is to be distinguished from a "development phase." By a development phase we mean a period of time of continuous development. In some cases the process and the phase coincide while in others the process is made up of a number of phases. In the latter case, the development phase ends when the price of A drops below the critical price. For example, in figure 11 if \( P_A \) dropped to $7 in period 2 this would be sufficient to choke off further development. But \( b_t \) is still positive and so if and when \( P_A \) subsequently rose above the critical price this would allow the development process to continue. Development
FIGURE 11
may also be choked off by increases in the critical price. We have already mentioned one reason for this occurring: the output of B is large relative to the total market. Four other reasons are:

(a) Rising input costs in the production of B;

(b) Attempts by the producers of A to differentiate their product from B;

(c) Attempts by the producers of A to place legal restraints on the sale of B (for example production quotas); and

(d) Increases in the rate of return from investments that compete with B for the entrepreneur's capital.

To show how this theory fares in practice let us now take a look at four substitutes that appear to have completed most of their development process. They are synthetic rubber, synthetic detergent, rayon, and margarine.

IV. FOUR ILLUSTRATIVE EXAMPLES

In this section an attempt is made to identify the periods when major steps forward were made in the development of four substitutes. A previous study by Schmookler [17] attempted to discern such periods of inventive activity by counting the number of patents issued each year. This is not a very satisfactory method to use here since it gives equal weight to all patents whether their development prospects are large or small. If one were to pursue such a method he should weight each by its likelihood of being acted upon and also by its expected impact were it to be used. This would be a monumental if not an impossible task, and will not be used here. Rather, an attempt is made to find periods where, (a) production of the substitute has been rapidly expanding,
(b) price of the substitute has dropped relative to that of the natural product, and (c) major technical breakthroughs have occurred.

For each example, we present a scenario of the substitute's development. First the history behind each substitute is presented. Then our dynamic model graphically depicts a scenario of development.

1. Synthetic rubber

Laboratory research in the nineteenth century produced a number of interesting materials with qualities similar to natural rubber. However, development of a useful syntetic rubber had to await the onset of World War I. Faced with a shortage of rubber at this time, Germany developed methyl rubber. During the final stages of the war, Germany was manufacturing some 300 tons per month. When the war ended however, all production and development ceased. The next phase of development began in 1925. In that year the price of natural rubber in England soared to 6/- per lb. where three years previously it had been 7 1/2 d. per lb. The cause of this was the "Stevenson Restriction Scheme" which restricted supply for the purpose of raising its price. This stimulated the German I. G. Farbenindustrie to renew its activity in the field of synthetic rubbers, one which it had dropped since 1918. It also stimulated chemical concerns such as DuPont to launch investigations into this area. However, shortly after the meteoric price rise in rubber, came an equally meteoric fall. This was due to the flooding of the market with rubber from Dutch growers in Sumatra who were outside this scheme. Despite the price fall the development of synthetic rubbers continued in Germany for a non-price reason: this was economic nationalism. Germany, after being severely hampered in World War I by a lack of rubber dearly wanted
to be self-sufficient in this commodity despite the higher price. In fact, Germany imposed a heavy import duty on natural rubber calculated to bring in $25 million that was to be spent on synthetic rubber development. The same reason spurred development in this field in Italy and Russia. In 1939 USSR production of synthetic rubber was 50,000 tons, German production was 20,000 tons while in USA only 3,000 tons was being produced.

It was not until the acute rubber shortage in World War II that the United States turned to synthetic rubber production on a large scale. By 1941 production had increased to 12,000 tons and by 1944 to over 1 million tons.

The price of synthetic rubber (neoprene) in 1931 was $1 per lb. However, by 1941 this had dropped to 65 cents per lb. as consumption and output increased. At this time natural rubber was 23 cents per lb. and production of synthetic rubber was still small-scale. As the scale of production grew the price of synthetic rubber fell. When the war finished the price of the most widely-used synthetic rubber was below the average price of natural rubber. It was also in some respects superior to natural rubber.

The superiority lay in both its consistently good quality (unlike natural rubber) and its stable price. The unstable price of natural rubber was equally unsatisfactory to the buyer when it was low as when it was high. This was evidenced following the collapse of the Stevenson Scheme when the major American rubber buyers were forced to buy rubber all the way down.
Synthetic rubber passed through three major development phases before it emerged as a major product in its own right after World War II. Our dynamic model suggests the following scenario for the development of synthetic rubber.

Prior to World War I we are at a point on line segment AB in figure 12: the price of natural rubber is not high enough to stimulate the development of substitutes.

With the advent of W.W. I and higher rubber demand we move out to point C. This stimulates the development of substitutes, resulting in a fall in PC (critical price) by the end of the war to OA'. However, the end of
war brings an even larger fall in PA (price of the natural product) so that we move to D and development is choked off.

The next phase of development coincides with the Stevenson Restriction Scheme. The rise in rubber price is shown by the move: D to E. The resulting development lowers PC in the move: E to F. Now, however, the action of Sumatran growers takes us to G and again development is choked off.

The third development phase coincides with World War II. We move out to H, development occurs and PC falls. Following the war, PA falls but this time not sufficiently to choke off development. Development may now proceed virtually unhindered by conditions in the market for the natural product and virtually independent of any price action that producers of the natural product may wish to take.

2. Synthetic detergents

Initial research into synthetic detergents was conducted by a Belgian, Reychler, as early as 1913. However, no serious development occurred until 1917. This was the time of the allied blockade of Germany during World War I. Germany was cut off from access to imported fats and oils. In an attempt to overcome the acute shortage that resulted, German chemists began to develop soap substitutes not requiring the use of fats. Despite the discovery of a detergent that could be made from coal tar the product could not be commercially developed before the war ended.

The next stage of development came in the mid-twenties. At this time, the textile industry was experiencing a boom. However, the large use of acidic solutions in the manufacture of textiles created problems
in the cleaning processes. Soap reacted with the acid to form a precipitate of fatty acid. The synthetic detergent, however, was not subject to this problem and so was in demand despite the fact that it was much more expensive than soap. Another advantage was the fact that synthetic detergent was not affected by hard water which was prevalent in much of the United States and Europe. Soap, on the other hand, declines a great deal in efficiency in hard water. As a result of these advantages Allied Chemical and Dye Corporation brought the German detergent to the United States. Shortly after a new cleaning agent was developed which was a derivative of petroleum.

Until World War II the principal use of synthetic detergents was in industry. Pre-war penetration of the household market was small, being confined to light laundering and dishwashing. There was no suitable substitute for soap in heavy laundering. The detergents that were available were disadvantaged by their high cost and the fact that housewives found them hard on the hands.

World War II, however, brought a soap shortage to the U.S. and caused a flood of synthetic detergents into the household market. In 1946 the development of a synthetic suitable for heavy domestic laundering received an added boost. The U.S. aviation fuel plants were converted to the production of tetrapropylene from which an excellent household detergent could be made.

Since 1946 the growth of the synthetic detergent industry in the U.S. has been very rapid. Sales increased from 125,000 tons in that year to 655,000 tons in 1950 to 1.2 million tons in 1955. Synthetic detergents thus captured the household market for heavy as well as light laundering
and dishwashing. Only bathroom soap has managed to keep its market intact. It is suggested a major reason for this is that cleansing ability is not so critical in this end-use as is softness on the skin.

Like synthetic rubber, synthetic detergents appear to have experienced a three phase development process. Our dynamic model suggests the scenario summarized in figure 13. (Unlike the previous example, only brief explanatory notes accompany this figure.)

**FIGURE 13**

First phase: A + B + C + D (World War I)
Second phase: D + E + F (textiles boom)
Third phase: F + G (World War II)

The first and third phases were definitely associated with periods of shortages of the natural product. The second phase is associated with
a high demand for an industrial cleanser. This period involved a high demand for soap but it also involved some dissatisfaction with the performance of the natural product in its industrial cleaning tasks. Both these factors contributed to the development of synthetic detergents in the second phase.

3. Margarine

During the mid-eighteen hundreds the huge population shifts in Europe from the country to the city, coupled with a series of poor seasons caused butter production to decline and prices to soar. At this time the far-off colonial dairy industries were still in their infancy and could not be counted on to relieve the situation. There was an urgent need for an acceptable butter substitute for the poorer classes. Napoleon III of France had a special interest in finding such a substitute to feed his army that was engaged in war at this time. In 1869 he offered a prize for a butter substitute that would be cheaper and would keep better than cow's butter. A Frenchman, Mége, patented margarine in that same year and collected the Emperor's prize. Further, he received financial assistance from the Government to open his first factory in 1871. Margarine sold for about two-thirds the price of butter and must have been tasty since the sophisticated French palate took to it very quickly. By 1873 the butter substitute was being produced in the United States. In 1875, cottonseed oil began to be used in the production of margarine since it provided a softer spread. In 1879 the Agricultural Chemist reported that "carefully prepared oleomargarine is superior to poor butter as regards taste, odor and healthfulness." The spread was colored to resemble butter with a vegetable substance, annatto. The new spread
was selling so well in the U.S. that it wasn't long before it ran into opposition from dairy producers who saw their livelihood threatened. Beginning in 1886, came a series of taxes and restrictions that severely curtailed the continued development of margarine in this country. A 10-cent Federal tax lifted the price of pre-colored margarine to that of butter, and heavy license fees for selling the pre-colored margarine led most retailers to not stock it. The next main phase of development came during World War I. The severe fats shortage drove up the price of butter from a pre-war average of 30 cents per lb. to over 60 cents per lb. in 1919. It was during this time that Proctor and Gamble introduced vegetable fat hydrogenation into the American margarine industry (though it had been used as early as 1906 in Europe). This meant that animal fats need no longer be used to obtain a solid product. Thus, the margarine industry could operate independently of the meat packer and dairy industries, and cheap cottonseed oil and coconut oil were extensively used. The third phase began in the mid-thirties amid the growing threat of another war (hence, expected shortages) and when there was mounting political pressure for the repeal of the restrictive laws governing the sale of margarine. In 1936 an important development was the introduction of the "votator". This enabled the appropriate oils and water to be combined in a far more stable chilled emulsion than was allowed by the earlier processes. Previously, margarine was very prone to deterioration as water separated from the fats in the margarine leaving the product with a wet appearance.

During World War II margarine made large market gains as a result of butter rationing. And after the war, the final blow to butter came
when a crisis in the dairy industry sent production costs soaring. In December 1947 the average retail price of butter reached $1 per lb. This directly led to the repeal of the Federal tax on pre-colored margarine, and hence to a continuation of the meteoric rise in the market share of margarine experienced during the war years. In subsequent years, state taxes and other restrictions have also been gradually lifted.

Our dynamic model suggests the scenario summarized in figure 14.

\[
\begin{align*}
\text{FIGURE 14} \\
\text{First phase: } & A \rightarrow B \rightarrow C \text{ (European butter shortage, mid-1800's)} \\
& C \rightarrow D \text{ (taxes and restrictions on margarine, late 1800's-early 1900's)} \\
\text{Second phase: } & D \rightarrow E \text{ (World War I, fats shortage)} \\
& E \rightarrow F \text{ (fat hydrogenation)} \\
& F \rightarrow G \text{ (end of fats shortage, depression)}
\end{align*}
\]
Third phase:  G → H  
(expectation of war and butter rationing, 1930's)

H → J  
(votator, 1936)

J → K  
(World War II, butter rationing; 1947, large increase in dairy production costs)

K → L → ... (end of Federal tax on margarine)

4. Rayon

The origins of rayon lay in attempts to produce an artificial silk. As far back as 1664, Robert Hooke, the English naturalist, suggested the possibility of artificially producing a fiber that resembled silk. Hooke realized the benefits that would accrue to the person who could duplicate the fiber used to make the garments of kings. However, it was not until the second half of the nineteenth century that research yielded the product that came to be known as "artificial silk" and then rayon. Silk was a perennially high-priced fiber and it is noteworthy that both rayon and nylon began their careers as "artificial silk". In addition however, it is suggested one important cause of the initial developments was an epidemic in the silkworms reared in France. From 1847 to 1875, this epidemic severely affected the industry. For example, in 1853, France produced 25 million cocoons but by 1865 this had fallen to 5 million. During this period, in 1855, a Swiss, Audemars, took out the first patent for artificial silk. However, the process was not successful in producing a thread of commercial value.

The next phase in the development of artificial silk was the period 1880 to 1900. During this period rayon evolved to become a successful commercial venture. Four different types of rayon were developed. The acknowledged pioneer was a Frenchman, Chardonnet. The expanding American demand for silk (especially from France) coupled with some important
technical breakthroughs appear to be the keys to the development in this period. The breakthroughs were in the development of spinnarets through which the filament yarn could be extruded. The early rayon industry owes much credit for this development to the electric lamp. The spinnaret was originally used in the production of filaments for the lamps. It was only a short step from there to its use in the production of filament yarn. In 1900, Chardonnet rayon was selling in London for 12/- per lb. while silk averaged 14/-.

This rayon was made of nitro-cellulose which had two major disadvantages despite its early success. The process was expensive and the product was very flammable. This led to the development of viscose rayon which came onto the market in 1905. Success was moderate and slowed up during World War I.

The next phase of development occurred during the 1920's. At this time there was a greatly increased demand for elegance in articles of wearing apparel. In particular, the change in ladies fashions to shorter dresses encouraged them to switch first from cotton to silk stockings. They then switched from silk to rayon stockings as the price of silk soared.

There was a rush of new firms into the rayon industry in this period which greatly increased the degree of competition. This led to international agreements in the late 1920's to eliminate wasteful competition but at the same time it gave an added impetus to the desire for innovation. In 1927, Courtaulds, the leading rayon firm, began producing acetate rayon. A principal advantage of this over viscose rayon was that it was a cheaper method of producing the fine yarns. The product also
was not shiny like viscose. As fashions changed from shiny to delustered material in the late twenties, so consumers switched back from viscose to silk, and then to acetate as that came on the market.

There were many other technical developments during the late twenties and early thirties that enabled prices to drop considerably. This price drop was necessary if the public was to accept the large increases in output. One such development, perhaps the most important, was staple fiber. This was rayon's answer to cotton and wool. It was first produced in 1918 by Courtaulds. However, it was of poor quality and could not compete with cotton and wool prices which at that time had slumped well below their wartime levels. Nothing more was heard of rayon staple until 1925 when wool and cotton prices were soaring. A new viscose staple fiber was introduced in that year, of far better quality than the original one. Its sales increased steadily until the depression when severe falls occurred in wool and cotton prices. In 1934 came the next phase of development for rayon staple. This was for two reasons. First, Courtaulds were able to acquire a much more efficient machine from an Italian firm and second, there was a growing threat of war which indicated a high expected demand.
FIGURE 15a: Rayon vs. silk

First phase:
A → B  (silkworm epidemic)
B → C → D (Audemar's "artificial silk")

Second phase:
D → E  (expanding silk demand in the U.S.)
E → F  (development of spinnarets, viscose rayon)

Third phase:
G → H  (heavy demand for silk stockings in 1920's)
H → J  (research on nylon)
J → K  (depression)
K +  (World War II encouraged the development of nylon and dacron as substitutes for rayon)

FIGURE 15b: Rayon vs. cotton

First phase:
A → B  (increased cotton demand in 1920's)
B → C  (development of new viscose staple fiber)
C → D  (depression)

Second phase:
D → E  (growing threat of war, 1930's)
E → F  (large-scale production, new efficient machines)
F +  (onset of war encouraged the development of a nylon substitute for rayon)
Throughout these four examples the development of substitutes occurred primarily in response to abnormally high prices for the natural product, particularly in the early stages of development. In so doing, they provide support for the theory presented earlier. In the next section is discussed an industry into which substitutes have only just begun to penetrate. The main purpose of this is to show how the development process was begun and how perhaps the industry may best combat it.

III. IMPLICATIONS FOR THE BEEF INDUSTRY

The size of the market for U.S. beef is enormous compared to those of other agricultural products. In 1972, U.S. cash receipts for cattle and calves was $18 billion which represented 30 percent of total cash receipts from agriculture. This suggests that the beef market is a tempting one for developers of substitutes. Then in 1973 beef prices jumped to an unprecedented high. In Omaha the average (choice) slaughter steer price was $43.89 per 100 lb., up 22% from the previous year. This sparked off the development and production of beef substitutes. In March 1973, a blend of soy ("extender") and ground beef was introduced into grocery stores. Gallimore [6] points out that the newly developed products "resemble ground beef in texture and maintain this texture through cooking. . .and when mixed with ground beef, take on the flavor and color of the ground beef."

In addition to these meat "extenders" came a flurry of simulated meats called "analogs." They appeared on the market in prepared foods such as "noodles stroganoff with beef-flavored vegetable protein chunks."
D. S. Greenberg [7] remarked on this: "the 'beef' pieces were small but could easily have passed for overcooked, heavily sauced bits of meat." It appears there is still a considerable quality difference between meat analogs and the real thing, and the cost of production is high because the scale of production is low. However, in 1973, with high prices for real meat, the major companies, General Mills and Miles Laboratories, rapidly expanded their research programs.

During the first half of 1974, the beef industry has gained a reprieve from the onslaught of substitutes. Beef prices have declined while the price of soybeans, an important ingredient in the substitutes, has risen. Local supermarkets have stopped offering the extended hamburger for sale and it appears that consumer interest in the meat analogs has waned. It is not known whether the major companies have since reassessed their research expenditures but it seems doubtful since the present decline in beef prices is generally expected to be only temporary.

Attempts by supporters of the beef industry to stem the tide of meat substitutes have included:

(a) Advertising the merits of the natural product
(b) Attacking the nutritional value of the substitute, and
(c) Insisting on labeling regulations that banned the use of words such as 'meat' to describe the substitute.

Of the three, only the first seems to hold promise in the event of a repetition of 1973 price behavior. The nutritional problem was largely beaten when the FDA certified the amino acid methionine for addition to textured vegetable protein. With this, the product is nutritionally
comparable to meat. The third attempt was designed to build a psychological barrier for consumers. But when the retail price of hamburger shot up to unprecedented highs, consumers were more than happy to take the lower-priced 'tasti-blend', 'juicy-blend' and 'better burger.' And once this initial skepticism is removed it is not likely to return.

Relying on advertising alone to carry the flag seems dangerous. For one thing, advancing technology in factory produced substitutes may some day leave the natural product with a net deficit as far as 'merits' go. Already the substitute manufacturers have gained two plus points with the "non-shrinking" and "low cholesterol" properties of their products.

A fourth possible way to stem the tide of meat substitutes is implied by the theory in this paper. That is: stabilize beef prices. The theory says that abnormally high prices for the natural product are a major incentive for the development of substitutes. But more than this, it says that following some development, the price of the natural product need not be so high to maintain further development. In the static model we saw this in the section on endogenous technological advance. In the dynamic model we saw this in the falling "critical price" (figures 9 and 11). Thus, if we can avoid abnormally high beef prices (by stabilization), this may considerably inhibit the development of substitutes. It is true, price stabilization has been advocated before but not, to my knowledge, for the purpose of inhibiting the development of substitutes. The main purpose of this paper has been to show how high prices for natural products are often associated with the irreversible development of substitutes. It follows that if their
prices were stabilized at the long run average, this would tend to
discourage the development of substitutes.

IV. SUMMARY

The development of substitutes poses an important problem for many
agricultural producers. It results in large scale resource reallocation
which can be very painful to producers of the natural product.

This paper has adapted conventional supply and demand theory to
explain why such development occurs. It was suggested that the complete
development process usually occurs over a number of phases. Initially,
we used a static framework to examine the cases where natural product
and substitute are of equal quality, and then where the latter is inferior
but improving with development. We subsequently discussed development
using a dynamic framework.

The main emphasis of the theory was on how development builds on
itself during a single phase and how it is irreversible between phases.
The latter it was suggested led to a ratchet-like decline in the critical
price as the price of the natural product fluctuated.

To support the contention that high prices for natural products are
indeed associated with the development of substitutes we looked at four
illustrative examples.

To halt the ratchet-like decline in the critical price it was
suggested the most effective method might be to stabilize the price of
the natural product. In the final section this suggestion is proffered
to the beef industry, an industry which is now ripe for the appearance
of substitutes.
V. Appendix

In figure 16 we wish to show that the vertical distance KQ is equal to PC(t + 1). For now, let us call this distance x. Then clearly NP also equals x, (where N is directly below our starting point M). In addition:

\[ LP = PA(t) \]
\[ MP = PC(t) \]
\[ \therefore LM = PA(t) - PC(t) \]
\[ \text{and } MN = PC(t) - x \]

Now LM + MN = KN (since the slope of LK is 1). Further, the slope of \( KM = \frac{b_t}{1 + b_t} \) (slope of the response function).

\[ \therefore \quad \frac{b_t}{1 + b_t} = \frac{MN}{KN} = \frac{PC(t) - x}{[PA(t) - PC(t)] + [PC(t) - x]} \]
\[ = \frac{PC(t) - x}{PA(t) - x} \]
\[ \therefore b_t[PA(t) - x] = (1 + b_t)[PC(t) - x] \]
\[ \therefore b_t PA(t) - b_t x = PC(t) - x + b_t PC(t) - b_t x \]
\[ \therefore x = (1 + b_t) PC(t) - b_t PA(t) \quad (i) \]

Our response function is:

\[ PC(t) - PC(t + 1) = b_t[PA(t) - PC(t)] \]
\[ \therefore PC(t + 1) = (1 + b_t) PC(t) - b_t PA(t) \quad (ii) \]
Comparing equations (i) and (ii), \( x = PC(t + 1) \), and we can see that the vertical distance KQ is equal to \( PC(t + 1) \).
VI. REFERENCES


[12] Knorr, K. E. Rubber after the War, War-Peace Pamphlets No. 4, Food Research Institute, Stanford University, 1944.


[20] Schwitzer, M. K. Margarine and Other Food Fats, Leonard Hill, 