I. Introduction.

Paul David has been a pioneer and innovator in so many directions that it is hard to keep track of them all. He has moved with ease from specific historical events and episodes to broad generalizations, both theoretical and empirical in nature. The diffusion of the reaper, with close attention to both social interaction and profitability, the need for purchasing power comparisons to replace foreign exchange rates in international welfare comparisons, the bias toward labor-saving innovation in explaining the evolution of the distribution of income, the interpretation of Habakkuk's thesis on American economic growth relative to England's, and, more recently, the comparative roles of science and technology in technical change are merely a few of the many contributions that Paul has made to the progress of economics and in particular to the analysis of technical change and its economic implications.

I want here to discuss one particular thesis that Paul has been strongly associated with, the idea of path dependence. This is a concept whose general meaning is fairly clear, though a precise definition is not so easy. Roughly speaking, it means that the long-term historical evolution of an economy (or any other system) depends on where it
started from or perhaps on some of the disturbances to the system during its history. The critical point is that the effect of these initial conditions or disturbances is essentially permanent; it does not gradually vanish with time.

Many economic theories do not show path dependence, e.g., the Solow growth Model, which predicts that all economies will converge to the same limiting capital-labor ratio even though they may start with very different ratios. It is a thesis of Paul David’s and others’ that real economies and parts of display path dependence. (I survey some of the history of this concept in Arrow [forthcoming]).

I do not think there is the slightest question that path dependence is a real phenomenon in economic history and development, as it is in biological evolution and in the history of political and social institutions. My purpose in this paper is to examine one claimed aspect about the explanation of path dependence, namely, that it arises as a result of increasing returns to scale in some relevant part of the economy. I will argue, by means of an example, that this is not so. Even with constant returns to scale and perfectly competitive equilibrium, path dependence is possible.
My suggestion is that it is the irreversibility of investment, not increasing returns, that is at the root of path dependence. Even though Paul David has emphasized the role of increasing returns in path dependence, my thesis is rooted in one of his papers, that on the failure of Great Britain to adopt the reaper (David [1971]). A study of his paper convinced me that the argument there did not depend in any way on increasing returns, and the model of this paper is loosely suggested by Paul's analysis.

II. The Formal Notion of Path Dependence.

A. Dynamic Systems in General: In this paper, I confine myself to deterministic dynamic systems. The evolution of the system is governed by the dynamic relations and the initial conditions. Suppose the system converges for all or some set of initial conditions. There is path dependence if the limit depends on the initial conditions, that is, if for different initial conditions, the system converges to different limits. The classic case is that of a drop of rain falling on a hill. Once landed, the water's flow is determined by the law of gravity and the particular topography. The rain will eventually go into a valley, but which valley depends on the point of initial contact with the ground.

B. Intertemporal Competitive Equilibrium: The particular dynamic system we
want to look at is intertemporal competitive equilibrium with an infinite horizon. The
technical framework was first formulated by Lindahl [1929] and subsequently by Hicks [1939]: the infinite-horizon version of a complete general
equilibrium was first studied by Bewley [1972]. It is now known that such models are
capable of extremely complex behavior, even chaotic solutions (Boldrin and Montrucchio
[1991]). It is perhaps therefore not surprising that path dependence is possible in
intertemporal equilibria.

Note that in competitive models, including intertemporal competitive models, we
are assuming concavity of the production functions and therefore rejecting increasing
returns.

III. Some Examples of Path Dependence

A. Path Dependence in General History: Historians, unlike economists,
have always been very prone to assume path dependence. Large consequences flow from
the idiosyncrasies of kings or other leaders (e.g., Henry VIII’s love life and the separation
of England from the Catholic Church). Blaise Pascal’s remark that the history of the
world would have been different if Cleopatra's nose were somewhat longer would be compatible with much historical analysis, though Pascal himself was no historian. I ran across the work of the great Dutch historian, Pieter Geyl [1961-4], on the separation of the Netherlands from Belgium. He argued that, to begin with, there was little difference between the northern and southern Netherlands, as they then were. Politically, they were a set of fiefs jointly under the effective sovereignty of the Spanish rulers. The resentment of Spanish rule was common. The religious division between Protestants and Catholics was about the same. However, as the Spanish sought to reestablish their control against the revolt, the convoluted sea coasts in the north provided greater refuge for the rebels. As the Spanish regain control in the south, Protestant refugees fled north, changing the religious balance. It was therefore because of geography that the north became the successful point of resistance and also the center of Protestantism, to the point that when independence was ceded to the north, the two areas became increasing divergent in religion and other social attitudes. Two hundred years later, the unity of the Netherlands was reestablished by the decision of the victors over Napoleon but could not be maintained with such divergence of religion and of national sentiment.

C. Veblen on German Economic Development: Thorstein Veblen's study of the rise of Germany as an industrial power [1915] is a prime example of path
dependence in economic history. Veblen’s rich analysis covers many aspects of German development and still repays study, but the particular point I want to emphasize is his explanation why Germany, a latecomer to industrial development compared with Great Britain, managed to surpass it. If we assume the two countries had access to the same technologies in, say, 1870, but Great Britain had accumulated more capital, we would expect according to simple models that Germany would gradually catch up with Great Britain but would always lag behind it. As generally perceived, though, the latecomer actually surpasses its rival.

Veblen’s view was that this was not an accident but in fact the natural result of being a follower. His argument was illustrated by the railroad systems and the corresponding freight-handling equipment at the ports. The British started with narrow-gauge railroads and built equipment to transfer freight from railroads to ships which matched. Later technological developments showed that standard-gauge railroads were more efficient. The Germans, building later, followed the late technology. The technological knowledge was equally available to the British, but since they would have had to replace both railways and freight-handling equipment, it was never worthwhile.
Hence, though both countries had access to the same technology and both were responding rationally, the Germans would develop higher productivity than the British.

In Veblen's account, there is path dependence because the future course of economic development is determined, even in the long run, by the initial capital configurations around 1870.

C. Urban Agglomeration: It is fairly common to observe that industrial and commercial activity is concentrated in cities rather than being spread out uniformly. Further, while some cities clearly have some special locational advantage, such as having a fine port, the locations of many do not. Many are located on plains where there seems no special advantage over many other sites. It is has long been a common argument, already found in Marshall [1920], that there are external economies which explain the structure of cities. This point of view was elaborated in the work on the economics of location; see Christaller [1966], Lösch [1954].

This insight can be given a dynamic interpretation. Given an initial unequal distribution of activity, the locations with denser activities have an advantage which
causes new activities to accumulate there, reinforcing the original inequality. This
process has been elaborated on by Arthur [].

B. Lock-in in Product Development: It has several times been claimed that the
specific characteristics of products can be historically determined. An outstanding
example is Paul David’s argument that the configuration of the typewriter keyboard was
determined by historical accident (David [1985]). In early typewriters, there were many
different assignments of letters to keys. The one now used had as its chief advantage not
speed but avoidance of jamming of keys. As the technology improved, this last property
became irrelevant. But in the meantime there was an accumulation of human capital
invested in learning the keyboard, an investment which made it uneconomic to switch to
a new keyboard designed for speed, even though such existed.

Arthur [1989] gave a general model in which development of a product in one
direction lowered the costs of continuing in that direction, even though an alternative path
would ultimately have yielded a greater product. He argued that the development of the
video-cassette recorder, in choosing between two rival formats, illustrated his model.
E. British Non-Use of the Mechanical Reaper: As I have indicated, interpreting David [1971] has led to this paper. Paul David sought to explain why the mechanical reaper, whose use was so widespread in the United States, was not used by British farmers. He argued that it was due to the plowing practices which were well adapted to an economy with hand reaping. The furrows were very deep, made more so by centuries of repeated plowing. As a result, the reaper would cut only the grain growing at the tops of the furrows, leaving the rest to be discarded or reaped by hand. Hence, there might be no gain to the farmers from switching to the reaper. Of course, the fields could be replowed to adapt to the mechanical reaper; but this would require a large capital cost which, it might be inferred, would not pay in subsequent labor saving. The American farmers, starting with virgin soil, had no such capital cost.

F. Is Path Dependence a Consequence of Increasing Returns?: Some of the examples given above certainly assume increasing returns. It seems to be assumed by Paul David, by Arthur, and, I suppose, by many others, that path dependence is intrinsically linked with increasing returns. Actually, the presence or absence of increasing returns in these examples is more subtle than might be thought. In the typewriter keyboard case, for example, the indivisibility (I take for granted that increasing returns always depends on the presence of an indivisibility) is in the acquisition of human capital (the skill of working at a particular keyboard). A similar
analysis can be made of the other cases. But in the last example, the non-use of the mechanical reaper in Great Britain, no increasing returns are present at all. The argument is just as valid on an arbitrarily small piece of land with an arbitrarily small amount of reaper-time.

There is one aspect that is common to all the examples: the durability of the capital. If plowing were renewed each year or if typists had to relearn their keyboards at short intervals, none of the lock-in, path-dependent character of economic history would be presence. I will present a simple model to show that even under all of the conditions which characterize competitive equilibrium, including constant returns to scale, path dependence is possible when there is an irreversible element to capital formation.

IV. An Illustrative Model

There are three production processes 0, 1, and 2, with successively increasing labor productivities. Process 0 requires no capital, processes 1 and 2 require specific kinds of capital (not shiftable across processes). Some capital of type 1 exists, none of type 2; process 2 has just been invented. Neither kind of capital depreciates. Each production process has fixed coefficients for labor and for capital. Both kind of capital are produced by labor alone. There is one consumer product, produced by the three
processes: utility is linear in the product. There is one consumer, who maximizes the integral (over infinite time) of discounted consumption, with force of discount (consumption interest rate) equal to \( \rho \).

Let \( N_i \) be the total amount of labor used in process \( i \). Let \( M_i \) be the amount of labor used to produce capital of type \( i \) (\( i = 1, 2 \)). Let \( V_i \) be the amount of labor needed to produce one unit of the consumer good by process \( i \) (\( i = 0, 1, 2 \)). Let \( u_i \) be the amount of labor needed to produce one unit of capital of type \( i \) (\( i = 1, 2 \)). Finally, let capital be so measured that in both processes 1 and 2, one unit of capital is needed to produce one unit of the consumption good.

Let \( K_i(t) \) be the amount of capital of type \( i \) at time \( t \). \( K_i(0) \) is some given positive number, \( K_i(0) = 0 \).

Given these conditions, an optimum path, which is also a competitive equilibrium in this simple one-person case, is determined. It will be shown that, for a large range of values of the parameters, the optimal path requires full utilization of the existing capital in process 1 forever; the remaining labor force is divided between building up capital in
process 2 and producing the consumption good from that process with whatever capital has been built up.

Let the total labor force be normalized at 1, so that, at any time $t$,

$$\sum_{i=0}^{2} N_i(t) + \sum_{i=1}^{2} M_i(t) = 1. \tag{1}$$

Total output of the consumption good is,

$$C(t) = \sum_{i=0}^{2} (v_i)^{-1} N_i(t). \tag{2}$$

The consumer (and the economy) maximizes the time integral of (2) discounted by $\exp(-\rho t)$, subject to (1), the capital accumulation equations,

$$\frac{dK_i}{dt} = (\mu_i)^{-1} M_i \ (i = 1, 2). \tag{3}$$

and the constraints in processes 1 and 2 that no more labor is applied than can be used in view of the amount of capital available,

$$N_i \leq v_i K_i. \tag{4}$$

Let $p_i (i = 1, 2)$ be the costate variables corresponding to the accumulation equations (3), $w$ the Lagrange parameter for the constraint (1), and $q_i (i = 1, 2)$ be the Lagrange variables.
for the constraints (4). Then the current-value Hamiltonian for the optimization problem is,

\[
H = \sum_{i=0}^{2} (v_i)^{-1} N_i + \sum_{i=1}^{2} p_i (\mu_i)^{-1} M_i + \sum_{i=1}^{2} q_i (v_i K_i - N_i) + w (1 - \sum_{i=0}^{2} N_i - \sum_{i=1}^{2} M_i).
\]

The evolution of the costate variables is then, given by,

\[
dp/dt = \rho p - v_i q_i \quad (i = 1, 2).
\] (5)

The first-order conditions for maximizing \( H \) with respect to the variables \( M_i, N_i \) are,

\[
v_i w \geq 1 \quad [N_i]
\] (6)

\[
v_i (q_i + w) \geq 1 \quad [N_i] \quad (i = 1, 2)
\] (7)

\[
p_i \leq \mu_i w \quad [M_i] \quad (i = 1, 2)
\] (8)

I will not go into a taxonomy of all the cases. I want to see if there is a range of parameters for which all new investment takes place in capital for process 2 but capital for process 1 is continued to be used fully. Since the system is linear (apart from corners), it turns out that we can confine ourselves to solutions in which the costate variables and Lagrange parameters are constant over time. We will assume this and demonstrate that such a solution exists.
From (5), stationarity of the dual variables implies that,

$$\rho p_i = q_i q_i (i = 1, 2).$$  \hspace{1cm} (9)

Suppose $M_i > 0$. Then equality holds in (8). In view of (9), this can be written,

$$q_i q_i = \rho \mu_i w.$$  \hspace{1cm} (10)

Then $q_i > 0$, so that the constraint (4) must hold with equality,

$$N_i = q_i K_i,$$  \hspace{1cm} (11)

and therefore $N_i > 0$. Hence, equality holds in (7). Together with (10), there are two equations in $q_i$ and $w$. In particular,

$$w = (\rho \mu_i + q_i)^{-1}.$$  \hspace{1cm} (12)

Clearly, we cannot have both $M_1 > 0$ and $M_2 > 0$, except by coincidence, for (12) would give us two different values for $w$. If we look for solutions with $M_2 > 0$, then we have in general $M_i = 0$.

If $M_2 > 0$, then $N_2 = v_2 K_2 > 0$, $M_1 = 0$ (except by chance), and

$$w = (\rho \mu_2 + v_2)^{-1}.$$  \hspace{1cm} (13)

The parameters have been chosen to satisfy (7-8) for $i = 2$. It must also be shown to be possible to satisfy condition (6) and conditions (7-8) for $i = 1$. If we substitute for $w$ from (13) into (7-8) with $i = 1$, we find after a little manipulation, the inequalities.
\[ 1 - v_1 (\rho \mu_2 + v_2) \leq v_1 q_i \leq \rho \mu_1 (\rho \mu_2 + v_2). \] (14)

Will there exist \( q_i > 0 \) satisfying (14)? It is necessary and sufficient that the first term be less than or equal to the third. This is equivalent to,

\[ v_1 - v_2 \geq \rho (\mu_2 - \mu_1). \] (15)

Since \( v_i > v_2 \) by assumption, (15) holds either if \( \mu_2 \leq \mu_1 \) (in which case process 2 dominates process 1 in both inputs) or if,

\[ \rho = (v_1 - v_2)/(\mu_2 - \mu_1). \] (16)

Finally, we have to ensure that (6) is satisfied; this guarantees that process 0 is not accepted. If again we substitute for \( w \) from (13), (6) is equivalent to,

\[ \rho \leq (v_0 - v_2)/\mu_2. \] (17)

If, then, the rate of force of discount is not too high, there will be a solution to the first-order conditions for which there is no investment in process 1 but the capital in both processes is fully utilized. The labor force is divided among three activities, producing in process 1 with the existing capital, producing in process 2 fully utilizing existing capital, and building up capital in process 2. The actual time path of these activities can be written easily but is irrelevant here.
The key point is that the capital and production level in process 1 remains always that in the original period. Thus even in this completely competitive model, with constant returns to scale, the equilibrium is path dependent. The path dependence arises from the irreversibility of the capital invested in process 1.

V. Comments

It seems to me to be an error to attribute path dependence to increasing returns. Rather, it seems to be due to irreversibility. All the examples which seem to imply the indispensability of increasing returns to path dependence also involve irreversibility, sometimes rather subtly, e.g., the irreversibility of human capital investment as in the typewriter keyboard example. This of course is not to deny the importance of increasing returns as a factor in economic analysis and in social policy formation.

REFERENCES


