Emphasising the institutional nature of the market in the above-discussed way also requires that we have to bring politics explicitly into the analysis of the market and to stop pretending that markets need to be, and can be, 'de-politicised'. Markets are in the end political constructs in the sense that they are defined by a range of formal and informal institutions that embody certain rights and obligations, whose legitimacy (and therefore whose contestability) is ultimately determined in the realm of politics. Consequently, institutional political economy adopts a 'political economy' approach not only in the analysis of the state but also in the analysis of the market. It emphasises the fundamentally political nature of the market and applies the political economy logic to the analysis of the market, and not just to the analysis of the state.

Ha-Joon Chang is Reader in the Political Economy of Development at the University of Cambridge. This chapter is abridged from Breaking the Mould: An Institutionalist Political Economy Alternative to the Neoliberal Theory of the Market and the State, *Cambridge Journal of Economics*, vo. 26, no. 5, 2002, pp. 539-59.

Economic Evolution and Cumulative Causation

George Argyrous

The notion of cumulative causation as a methodological principle for guiding economic analysis was first developed by Thorstein Veblen, one of the founders of the institutionalist movement. Essentially, cumulative causation explains the emergence of mass production and how it spreads throughout the system by forces of its own making. Veblen regarded cumulative causation a vehicle for transforming economics into an evolutionary science, as opposed to the static and mechanistic character it had taken under neoclassical economics. Since Veblen's time the concept of cumulative causation has itself evolved, and now offers a distinct methodology to that used by neoclassical economics.

Neoclassical economics is based on three assumptions. These assumptions are that tastes and preferences of consumers are fixed and unlimited; that the resources available to meet these wants are limited; and that the technology available to transform resources into goods that can satisfy some of these unlimited wants is given. The economic problem is thereby a static one: how to satisfy these unlimited wants with the limited resources available. To solve this problem another assumption is added regarding the strategy economic agents follow in making choices between these unlimited wants. This is the assumption of rational maximization. The details of these assumptions can be found in any standard micro textbook; the important point for the following discussion is the teleological nature of neoclassical methodology. Causality runs in one direction from the assumptions to the variables to be explained (prices, wages, profit, employment, etc.). Once we know the initial conditions the final resting point – equilibrium – is predetermined.

Cumulative causation is based on a different methodology. Rather than unidirectional causality from independent to dependent variables, each variable interacts with the others in a mutually dependent way. Thus tastes and preferences, technology, and available resources change during the course of economic growth. Moreover, the principle of rational maximization as an assumption about behavior is substantially qualified, so that the behavioral rules which agents follow (traditions, social norms, blind habit, rules of thumb, etc.) emerge out of the process of

adaptation and evolution to a changing environment, and which in turn affect the environment. Rational maximization may emerge as one of these decision rules – it certainly is relevant in an analysis of the imperatives on business to cut costs – but its relevance is contingent on the context.

The division of labor and the specialization of industry

Even though the model of cumulative causation is one in which 'everything depends on everything else', the foundations of the technological system are given central importance in most discussions. All the key figures who have explicitly developed the notion of cumulative causation begin with the effect that the application of fixed capital and the division of labor has on capitalist growth. The shift from craft production to mass production fundamentally altered the way economic growth occurs. The nature of this shift in production technology is discussed in the work of Allyn Young (1928) and Nicholas Kaldor (1966), who drew on the inspiration of Adam Smith. In the *Wealth of Nations* Smith argued that the division of labor is limited by the extent of the market. By this he meant that the application of heavy machinery and the breakdown of the production process into its component parts is limited by the ability to sell the larger output that such technology generates. With a larger market to cater to, economies of scale can be realized through the division of labor, and thereby bring about an increase in productivity and the social surplus. If this surplus is consumed productively it will bring about further expansion of the market, and thereby encourage the further application of heavy machinery and division of labor. The circular character of these induced changes means that in fact the division of labor is limited by the division of labor!

Adam Smith illustrated this process by analyzing the internal restructuring of the famous pin factory. Young and Kaldor took this one step further and argued that the specialization of tasks and the decomposition of the production process into sequential stages causes a vertical splintering of industry: each layer produces an intermediate product which becomes an input to the next stage in the production process. Each 'layer' of firms produces one component of the final commodity, which ultimately comes together in the final assembly stage. Pasinetti (1981) refers to such a network as a "vertically integrated industry", which is a far cry from the craft shop which would turn raw outputs into final product, and along the way make any machinery and equipment under the one roof.

In short, therefore, a number of vertically integrated firms insert themselves between raw materials and the final product – a process which we may call vertical specialization, and which raises productivity through learning-by-doing and dynamic returns to scale. The application of mass production technology progressively brings down unit prices as productivity grows, which causes the market to expand and open up the field for further extension of mass production within that industry.

This process of vertical specialization, if pushed far enough, leads to the establishment of a layer of firms which focuses specifically on the production of capital goods. Of particular importance is the emergence of a machine tool industry which produces equipment such as lathes, grinders, and milling machines. These pieces of equipment, as Marx pointed out, take over the functions that were once performed by skilled artisans, and which are only viable when substantial economies of scale have been reached. Since individual machines can be used to produce thousands of units of final product, the market for the latter has to expand considerably before a market for specialized equipment emerges. They are, in short, the technology upon which mass production is based and are the means by which thousands of units of identical, standardized products are produced. These machines are used to produce final or intermediate products, or indeed to produce more machines.

Rosenberg (1976) has discussed the way in which the formation of a capital goods sector, and

a machine tool industry specifically, gives the evolution of capitalism an added mechanism of cumulative change. This is a learning-by-using process, whereby the users of capital goods identify problems and limitations with existing designs of equipment and relate this information back to the equipment producers. This poses a technological puzzle which equipment producers solve by modifying and adapting successive existing designs. This process feeds on itself as each solution tends to create a new problem – the advent of high speed metal alloys increased the speed of many machine tools, but this required major structural changes in the design of the machines to deal with the vibrations caused by the higher speeds. The result is a technological disequilibrium where new, unforeseen problems emerge out of solutions to old ones.

This incremental development, according to Rosenberg, follows a "compulsive sequence" of problems-solutions-new problems, and is given a specific direction by "inducement mechanisms" and "focusing devices" which set this compulsive sequence in train. These factors ensure that the development path is *historically* conditioned. For example, Rosenberg, citing Marx, points out how disputes between capitalists and workers in Manchester in 1825 directly led to the invention of the self-acting mule to allow for greater control of the labor process. Another common source of direction is provided by wars which cause a major disruption to critical supplies. The implication is that historical accidents will not be 'washed-away' by supposedly permanent and systematic forces that direct the economy toward a long run equilibrium: "the mere cessation of interference will not leave the outcome the same as if no interference had taken place" (Veblen 1919: 116). In modern terminology, cumulative causation is *path-dependent*, since the specific sequence of industries that emerge and the specific nature of the technologies they adopt will usually be affected by historically contingent circumstances that have long-lasting effects. The history of the QWERTY keyboard (discussed in an earlier chapter) illustrates this point.

Cumulative causation across industries

The model of cumulative causation has thus far looked at the way in which the mass production of a particular commodity establishes a process of expansion for the layers of firms that are vertically integrated in its production. As this process takes hold in one industry, though, it then sparks the division of labor in industries producing entirely distinct commodities. The mass production of cars, for example, affects, and is affected by, the same process occurring in the production of shoes and washing machines, so that the production of various commodities becomes interconnected. There are three related transfers which release cumulative forces of expansion across industries: transfers of technological knowledge; transfers of organizational knowledge; and transfers of wage income paid in one industry as demand for the product of other industries.

Transfers of technology

Mass production is built on a set of core, and fairly generic, industrial processes, such as grinding, milling, and planing. All manufactured commodities will involve the application of some, if not all, of these activities. The use of a common set of basic industrial processes means that a solution to a technical problem in one industry can be subsequently used in the production of completely different commodities to solve similar production problems. However, the requirements from industry to industry will not be the exactly same, so that some adaptation of technology must take place. Technology, in other words, has a public and a private dimension (Nelson 1993). In so far as machines are designed for a specific production line the technology is appropriated privately by the user of the machine. However, in so far as it is the application of general engineering and design principles it has a public dimension, and which gives it a potentially wider sphere of application.

Rosenberg illustrates this with the example of the stocking lathe for the shaping of gunstocks

which was then applied in the production of hat blocks, handles, wheel-spokes, sculptured busts, oars, and shoe lasts. Moreover, as we noted above, an improvement in any one aspect of technology often disrupts other parts of the system so that new problems arise: "single improvements tend to *create* their own future problems, which compel further modification and revision" (Rosenberg 1976: 29). So there is a leapfrogging process whereby problems are generated and solved, and the solutions are diffused into sectors of the economy other than where they originated, creating, in turn, new production problems to solve.

The other key institution involved in the transfer of technology from one industry to another is the modern research and development division of the large corporation. The pre-eminence of production technology under conditions of mass production, the demands on energy production and use, the importance on fundamental scientific advance render R&D of critical significance. Through the desire of the corporation to take advantage of a new development, scientific breakthroughs quickly gain a wide application. The distinction between knowledge as a private and specific application of generic and public knowledge again feeds the cumulative process. There is a private incentive to apply and extend innovations in order to extract windfall profits and recoup the enormous costs involved in R&D, and this incentive is heightened by the fact that the public nature of the underlying technology could allow competitors to appropriate the returns instead (Nelson 1993: 15-17).

Transfers of organizational knowledge

The transfer of mass production technology involves more than just technical know-how. It calls forth an entire system of business organization: the rise of large industrial corporations. The vertical fragmentation of production brought about by the division of labor and economies of scale requires an immense amount of planning and coordination. Unlike the small craft shop where a commodity was made from scratch under the one roof, mass production splinters productive activity across thousands of separate units. Yet each of these units must coordinate with the others, otherwise the final commodity will not materialize: the smallest discrepancy between screw sizes and screw holes in a panel, for example, and the car will literally not come together. The maintenance of strict standards, the coordination of production runs and of delivery schedules, the regulation of stocks and supplies, and the monitoring of quality control become very large managerial problems which require skills of their own that come to reside, as Galbraith (1971) points out, in the large bureaucratic technostructure of the modern corporation.

Chandler (1990; 1992) has traced the historical evolution of the modern industrial corporation and the way it has facilitated the spread of mass production across industries. He argues that when an industry begins to realize economies of scale, firms who are the 'first-movers' into that industry are able to dominate it for years to come, using their accumulated managerial expertise to keep late-comers out. During a critical period in the late 1800s when mass production technology was emerging on a wide scale in the US, these first-mover firms were able to invest in the three key areas of production, distribution, and management, and thereby 'take hold' of the market. Once established within an industry these large corporations then began to realize economies of *scope*. These are the gains that can be made by using the organizational expertise developed in one market to enter into other markets and thereby continue growth and dominance. The managerial skills that form the technostructure have a generic quality which allows them to be applied across a wide variety of activities. Thus individual corporations provided the institutional framework within which mass production spread cumulatively across sectors. They have an incentive to do this because diversification spreads risk so that problems arising in one market do not undermine the whole company.

The transfer of demand

These two transfer processes, that of production technology and of organizational structure, are not sufficient by themselves to explain the cumulative spread of mass production across industries. These systems only become viable if there is a mass market to cater to, so some explanation of the way in which mass markets grow and in turn feed these other transfers is needed. Craft production is geared toward custom-ordering in which products are 'tailor-made' to the needs of individual orders placed in advance of production. Mass production is based on the manufacture of thousands of units of standardized, identical products, which are produced in anticipation of a large homogenous market. The question is how does one type of market eliminate the other?

It is at this point that the economic sphere of life and broader social processes are most closely tied. The conception of growth as a holistic process in which mutually reinforcing feedbacks between the economic sphere and cultural and political institutions bring about cumulative causation has been most emphatically argued by Gunnar Myrdal (1974: 735-6). According to Myrdal, the problem is not dissected into an economic component and a social component, but rather treated as an indivisible whole.

Central to the explanation of the rise of mass markets is the change that occurs to the character of the household. Households have always had to balance the amount of labor available to them to devote to domestic production, and the amount to be sold in return for paid income. In the pre-industrial era, most families satisfied the bulk of their consumption needs within the home, using any income generated outside domestic production to buy raw materials such as cloth and seed and applying domestic labor to transform them into consumable items. However, this basis for household organization gives way to households in which most labor power is sold for wages, and these wages are used to purchase commercially produced final and intermediate goods. With the formation of an industrial working class, less labor is devoted to domestic production and households become increasingly reliant on produced commodities to satisfy their consumption needs. The household is transformed from being a production unit to a consumption unit and a supplier of wage labor.

How does this actually happen? As labor is drawn in from the countryside when an industry expands, these workers will spend their wage income purchasing goods that they no longer have time to produce in the home, thereby creating markets for other industries. As markets for these other industries expand, the division of labor in these industries takes off to realize economies of scale, and labor drawn into these industries will further expand the market for other wage goods.

This change in the structure and function of the household *vis-a-vis* the market works itself out slowly. Households do not instantly change from being large extended families living in small towns to urban, nuclear families, nor are all production activities transferred to the mark*et al.* the same time. The stable consumption pattern exhibited by the Engel Curve suggests that some goods are transferred to the market relatively early as income expands, with others entering the bundle of wage goods as demand expands further. The demand for each commodity follows an S-shaped path of growth: the market expands rapidly in initial phases but eventually reaches saturation. The market is then geared to replacement purchases. But as long as incomes are growing, the decline in one set of markets will be offset by an increase in others. Slowly more commodities produced under industrial processes enter into the consumption patterns of households, replacing goods produced within the household itself. This approach to consumption provides a sociological basis for the explanation of consumer demand that takes into account learning processes and emulation between groups of consumers as their incomes increase (rather than following the neoclassical notion of rational choice at a *given* income level) (Pasinetti 1981: 69). An example of such a sociological theory of consumption was provided by James Duesenberry (1967) and his 'relative income hypothesis' which itself built on the work of Veblen.

Picturing the development of the economy in this way as a holistic process of qualitative change affects the method of economic analysis. Myrdal, for example, originally thought that the patterns of causal relations that generate cumulative causation could be "given in the form of an interconnected series of quantitative equations" and this would provide a "truly scientific solution" (1944: 1069) to the problem. However, he later substantially qualified the argument for a quantitative approach (1968: 1866-70), and finally rejected such a position altogether, arguing that the feedback mechanisms were essentially qualitative in character, and therefore not reducible to a set of equations. The "coefficients of interrelation among the various conditions in circular causation are ordinarily not known with quantitative precision" (1978: 774). Instead, detailed historical analysis of the particularity and peculiarities of individual industries and countries is needed. Certain aspects of this process might be amenable to quantitative measurement and formal modelling (especially drawing on evolutionary game theory, time series analysis, and systems theory), but the whole story can never be told this way. Armchair theorizing won't go very far: the methods of anthropologists, sociologists and economic historians (e.g. fieldwork and historical case studies) become relevant to economists as well.

The arrangement of industries around a core sector of capital goods producers presented here may be misleading in one important respect. It may imply that any given industry is as important as any other. However, this is not the case. The three factors we noted which connect change in one industry with change in another, also allow us to construct a hierarchy of industries according to their capacity to spur the cumulative process. The most important for twentieth century development is clearly the automobile. The technical complexity and the ensuing technical puzzles involved in mass producing cars generated many of the major improvements in equipment and machinery, which then fed into other sectors; the firms which dominated the auto industry such as GM and Ford were able to use their organizational strength to move into other sectors such as aeroplane manufacture; and no other single product has so affected the patterns of social organization. The facility that the automobile has given to the process of urbanization and the unification of previously disparate markets has caused the demand for innumerable other products to expand.

Limits to cumulative causation

This discussion of cumulative causation may give the impression that once set in motion, the growth process spirals upward with a vigorous energy of its own making. However, such an impression is misleading because there are factors which retard the process and which in fact can turn a virtuous cycle of growth into a vicious cycle of decline. The first limitation arises from the mutual dependence of variables on each other, which is the very basis of cumulative causation. An example of this arose in the development of the aircraft industry (Holley 1964: 27):

> Low-priced airplanes waited upon the introduction of production techniques in the industry, but high-volume production could be justified only by a mass market, which waited upon low-priced airplanes. Until some escape from this circle could be found, true mass production in the aircraft industry would remain out of reach.

Where everything, in a sense, depends on everything else, there may be system-wide inertia and overall stagnation that may be difficult to overcome on a sectoral basis. Thus some system-wide form of regulation is required, a function usually undertaken by the state. The state can operate as a circuit breaker if such a problem arises. By coordinating the integration of the various components, and by initiating growth in key sectors which trigger expansion elsewhere the state can set a virtuous cycle in motion.

Even if some exogenous nudge has been given to the system so that it begins to cumulate under forces of its own making, such cumulation cannot occur at too rapid a pace. The second force of inertia is due to the *sequential* nature of the causal relations, such that each variable in the sequence can only change within the limits set by the previous changes in the other variables. For example, new technology will raise productivity and lower the real price of the commodity produced. However, it may take time for consumers to incorporate this new commodity into their consumption patterns. It is like trying to walk with shoe laces tied together: the left foot can only step so far in front of the right, and the position of the right foot is itself determined by previous movements of the left. Over time great distances can be travelled, but obviously each foot cannot go too fast without causing the whole body to topple over. Cumulative causation is even more complicated because there are many feet tied together in all sorts of directions.

A particularly important step in this causal sequence, which can limit overall expansion, is the capacity of the machine tool sector. Its pivotal role within the input-output matrix is discussed above, and as Lowe argues this "strategic position" (1976: 30) imposes a constraint on the entire economy:

> One need only to consider an increase in the aggregate demand for coal ... Then we see at once that the critical bottleneck 'in the hierarchy of production' arises in the machine tool stage and that only after capacity has been increased there, can the output of ore, steel, extractive machinery and, finally, coal be increased (1976: 34, n.6).

The bottleneck is compounded though by the fact that capacity in the machine tool industry is directly limited by its own capacity. This is because machine tools are key elements in their own production. The expansion of capacity in other industries necessitates a further expansion in the production of tools and equipment, but this then requires an expansion of tools and equipment!

The third force of inertia relates to the fact that cumulative causation is inherently a process of social and institutional change and such changes do not occur quickly. Institutions and social practices operate on the basis of customs, traditions, and habits which are deeply ingrained and only slowly abandoned and replaced by others. For example, the transfer of productive activity from the household to the family in any given country takes over a century to complete. Households have to learn knew patterns of behavior and new forms of social interaction, largely through emulation; corporations have to adjust their routines and structures (Nelson and Winter 1982). In other words, the involvement of social institutions weakens the coefficients of interrelations, giving the system a level of stability that prevents cumulative growth or decline from accelerating too rapidly: "But certainly the main resistance to change in the social system stems from attitudes and institutions. They are part of an inherited culture and are not easily or rapidly moved in either direction" (Myrdal 1968: 873).

The fourth force of inertia relates to the discrepancy that can arise between individual actions and their collective social outcome. People act with certain objectives in mind following various rules of behavior, but in a social network these objectives may not be realized, so that the goal or the means to attain them need to be adjusted. An unintended outcome at the social level will cause the original basis for decision-making to change. This process of adjustment to environmental changes brought about by past actions is the very driving force of cumulative causation. It is why learning features so prominently in the model: there are no absolute rules to guide behavior at all times, and given the contingent, historical and limited nature of information available, consumers and producers adapt through learning.

In other instances, however, this discrepancy can also lead to stagnation. Of particular importance is the fact that each firm has an incentive to innovate by cost-cutting, especially by replacing workers with machines. For each firm, it essential that they try to raise productivity and to lower costs, but the aggregate effect of this on effective demand is to undermine markets which

are the precondition for such innovation. Thus mass production has an in-built tendency to stagnation that can overwhelm the forces that are also present and that facilitate growth and expansion.

A fifth limitation on the cumulative growth process is the finite nature of domestic markets. The satiation of domestic markets for consumer goods may cause the cumulative process to lose steam. Therefore, it is imperative that foreign markets be tapped if growth is to continue, once production for domestic markets has reached a saturation level (Kaldor 1966). This may be a fairly natural progression if the economies of scale realized in meeting home markets are sufficient to allow easy penetration of overseas markets. However, this will not always be the case, so that domestic markets are saturated before sufficient progress has occurred to make goods internationally competitive.

Moreover, other countries will be building up production for home markets as a basis for exporting, so that not all countries can pursue this strategy at once: there are necessarily winners and losers. The determination of winners and losers is not based on any inherent comparative advantage, but rather on the ability of any given economy to facilitate the forces which generate improvements in production. Advantages are created as a result of historical development rather than endowed by nature. And small advantages that one nation has over another at an early phase of development may accumulate over time so that it becomes a very wide discrepancy later, with one country on a virtuous cycle of expansion and the other on a vicious cycle of decline. A similar approach can be taken to explain differences between regions within a country in terms of differential growth rates. Growth rates do not converge across countries or regions, but rather become more pronounced.

A final limitation on the endogenous process of expansion has been discussed in detail by Setterfield (1997; 2001). Setterfield argues that the very process that drives the cumulative process of growth creates, as we discussed above, a high-degree of interrelatedness among the parts of the economics system. But this interrelatedness can 'lock-in' the system to a technology that becomes outdated and uncompettive with other systems not yet locked-in in this way. The mature system cannot 'lump tracks' easily and adopt fundamentally new technologies, precisely because it is a system and needs to adapt as a *system* rather than as individual parts. This is similar to the phenomenon Veblen described as the 'penalty of being first'. This clearly raises the necessity for such system transformations to be coordinated, probably by the state, so that once advanced economies do not find themselves becoming industrial backwaters.

Conclusion

The model of cumulative causation presented here provides a much greater scope for the government to alter the trajectory of industrial development in a positive way than in the neoclassical approach. A country is not restricted by its natural endowment of resources and comparative advantage – it has greater discretion to follow alternative development paths. Comparative advantage (as discussed by West in his chapter below) is a static concept relating to the relative status of countries at any given point in time. This state of affairs, though, is the outcome of an historical process that can take alternative paths. However, to paraphrase Marx, while nations do make their own history, they are not free to do this at will. The possibilities that are open for policy are limited at any point in time by the historically given conditions in which a nation finds itself. The conception of the economy as a set of mutually dependent institutions means that any aspect of it can only be altered within the limits set by other aspects. It is fundamentally a story of *incremental change*.

George Argyrous is Senior Lecturer, University of New South Wales.