Increasing Returns, Path Dependence, and the Study of Politics

Paul Pierson


Stable URL: http://links.jstor.org/sici?sici=0003-0554%28200006%2994%3A2%3C251%3AIRPDAT%3E2.0.CO%3B2-Z

The American Political Science Review is currently published by American Political Science Association.
Increasing Returns, Path Dependence, and the Study of Politics

PAUL PIERSON  Harvard University

It is increasingly common for social scientists to describe political processes as "path dependent." The concept, however, is often employed without careful elaboration. This article conceptualizes path dependence as a social process grounded in a dynamic of "increasing returns." Reviewing recent literature in economics and suggesting extensions to the world of politics, the article demonstrates that increasing returns processes are likely to be prevalent, and that good analytical foundations exist for exploring their causes and consequences. The investigation of increasing returns can provide a more rigorous framework for developing some of the key claims of recent scholarship in historical institutionalism: Specific patterns of timing and sequence matter; a wide range of social outcomes may be possible; large consequences may result from relatively small or contingent events; particular courses of action, once introduced, can be almost impossible to reverse; and consequently, political development is punctuated by critical moments or junctures that shape the basic contours of social life.

It is increasingly common for social scientists to describe political processes as "path dependent." Claims of path dependence have figured in both classic works of comparative politics, such as Lipset and Rokkan's (1967) analysis of European party systems, and more recent analyses on topics such as labor incorporation in Latin America (Collier and Collier 1991), the outcome of state-building processes in Europe (Ertman 1996), and the comparative development of health care systems (Hacker 1998). The notion of path dependence is generally used to support a few key claims: Specific patterns of timing and sequence matter; starting from similar conditions, a wide range of social outcomes may be possible; large consequences may result from relatively "small" or contingent events; particular courses of action, once introduced, can be virtually impossible to reverse; and consequently, political development is often punctuated by critical moments or junctures that shape the basic contours of social life (Collier and Collier 1991; Ikenberry 1994; Krasner 1989). All these features stand in sharp contrast to prominent modes of argument and explanation in political science, which attribute "large" outcomes to "large" causes and emphasize the prevalence of unique, predictable political outcomes, the irrelevance of timing and sequence, and the capacity of rational actors to design and implement optimal solutions (given their resources and constraints) to the problems that confront them. If path dependence arguments are indeed appropriate in substantial areas of political life, they will shake many subfields of political inquiry. This essay argues that they are.

The analysis begins with a general discussion of path dependence that seeks to clarify some important ambiguities surrounding the concept. I then outline and investigate the distinctive characteristics of social processes subject to what economists call "increasing returns," which could also be described as self-reinforcing or positive feedback processes. For some theorists, increasing returns are the source of path dependence; for others, they typify only one form of path dependence. The focus here is on increasing returns processes, both because they are of great social significance and because (in contrast to broader conceptions of path dependence) social scientists are beginning to develop rigorous arguments about the causes and consequences of increasing returns. Increasing returns dynamics capture two key elements central to most analysts' intuitive sense of path dependence. First, they pinpoint how the costs of switching from one alternative to another will, in certain social contexts, increase markedly over time. Second, and related, they draw attention to issues of timing and sequence, distinguishing formative moments or conjunctures from the periods that reinforce divergent paths. In an increasing returns process, it is not only a question of what happens but also of when it happens. Issues of temporality are at the heart of the analysis.

The following section reviews the development of increasing returns arguments in the social science discipline in which they have received the greatest attention: economics. This review suggests the wide sweep of potential applications, even in a field that might be expected to be hostile to the idea. More important, these economic applications provide the most analytically developed discussions of increasing returns. Economists not only have clarified the principal implications of path dependence but also have identified many of the specific aspects of a particular social environment that generate such processes.

The discussion of economics prepares the way for an exploration of the distinctive characteristics of politics.
Rather than simply apply extant arguments in economics to political phenomena, we need to consider the features of the political world that require modifications in the use of path dependence claims. I will demonstrate that increasing returns arguments are at least as relevant to an understanding of politics as they are in other areas of the social sciences. Indeed, factors such as the prominence of collective activity in politics, the central role of formal, change-resistant institutions, the possibilities for employing political authority to magnify power asymmetries, and the great ambiguity of many political processes and outcomes make this a domain of social life that is especially prone to increasing returns processes.

The final section considers what these arguments can contribute to political analysis. They provide an important caution against a too easy conclusion of the inevitability, "naturalness," or functionality of observed outcomes. Given the ubiquity of claims about efficient or functional elements in politics, this alone would be an important corrective. More significant, increasing returns arguments justify efforts to stretch the temporal horizons of political analysis. They can redirect the questions political scientists ask, which will contribute to a richer appreciation of the centrality of historical processes in generating variation in political life. They can also direct attention toward particular variables and suggest promising hypotheses about the sources of both political stability and political change in certain common political settings. For instance, increasing returns arguments highlight the need to consider hypotheses based on temporal ordering—the possibility that the particular sequencing of events or processes may be a key part of the explanation for divergent outcomes. Finally, grasping the implications of widespread path dependence can help orient political scientists to a realistic, which is to say modest, set of aspirations regarding the possibilities for achieving parsimony and predictability in the study of politics.

PATH DEPENDENCE AND INCREASING RETURNS

Analysts are increasingly inclined to invoke the concept of path dependence, but clear definitions are rare. In practice, usage tends to fluctuate between a broader and narrower conception. In the broader version, path dependence refers to the causal relevance of preceding stages in a temporal sequence. William Sewell (1996, 262–3), for instance, suggests path dependence means "that what happened at an earlier point in time will affect the possible outcomes of a sequence of events occurring at a later point in time." This usage may entail only the loose and not very helpful assertion that "history matters," although it may also be presented with more rigor. To take an example offered by Andrew Abbott (1983, 131), the contemporary significance of a group's organization of 40% of its potential membership depends greatly on whether that membership used to be 100% or 10%. Note that Sewell's definition involves no necessary suggestion that a particular path is difficult to exit. Rather, the claim is that we cannot understand the significance of a particular social variable without understanding "how it got there"—the path it took. Previous events in a sequence influence outcomes and trajectories but not necessarily by inducing further movement in the same direction. Indeed, the path may matter precisely because it tends to provoke a reaction in some other direction.

An alternative, narrower definition has been suggested by Margaret Levi (1997, 28):

Path dependence has to mean, if it is to mean anything, that once a country or region has started down a track, the costs of reversal are very high. There will be other choice points, but the entrenchments of certain institutional arrangements obstruct an easy reversal of the initial choice. Perhaps the better metaphor is a tree, rather than a path. From the same trunk, there are many different branches and smaller branches. Although it is possible to turn around or to clamber from one to the other—and essential if the chosen branch dies—the branch on which a climber begins is the one she tends to follow.

This conception of path dependence, in which preceding steps in a particular direction induce further movement in the same direction, is well captured by the idea of increasing returns. In an increasing returns process, the probability of further steps along the same path increases with each move down that path. This is because the relative benefits of the current activity compared with other possible options increase over time. To put it a different way, the costs of exit—from switching to some previously plausible alternative—rise. Increasing returns processes can also be described as self-reinforcing or positive feedback processes.

We face, then, a choice between a broader and a narrower conception of path dependence. Frequently, authors are not explicit about which of these two meanings they intend. To establish greater clarity, this essay employs the term in the narrower sense, that is, social processes that exhibit increasing returns. The fuzziness that has marked the use of this concept in social science suggests that the greater range offered by the broader definition has come at a high price in analytical clarity. Path dependence has been a victim of what Sartori (1970) called concept stretching. Different types of temporally linked sequences are generated in different ways and have different implications (Abbott 1983, 1990; Pierson n.d.b). These distinctive kinds of social processes, which have been bundled together, must be disentangled and systematically explored. Limiting the concept of path dependence to self-reinforcing processes in no way precludes the investigation of other ways in which sequences can matter in explaining social outcomes. It does encourage clear argument about distinct claims.1

1 Another strategy would be to use a broader conception of path dependence and then distinguish subtypes, exploring how these are generated in different ways, with different consequences (Mahoney n.d.). Because path dependence is often invoked without further clarification, however, this option seems more problematic. Ultimately, such definitional disputes remain intractable. What is critical is that researchers be clear and consistent about what they mean when they employ the concept and recognize the importance of differentiating among distinct processes.
However such issues of concept formation are ultimately resolved, there are two compelling reasons for focusing special attention on processes that exhibit increasing returns. First, such processes characterize many important parts of the social world. Second, social scientists are developing theory that makes the investigation of the causes and consequences of increasing returns a particularly promising area of inquiry. These two reasons are developed in later sections of this essay.

The basic logic of increasing returns processes can be captured in a simple mathematical illustration. Imagine a very large urn containing two balls, one black, one red. Remove one ball, and then return it to the urn, accompanied by an additional ball of the same color. Repeat this process until the urn fills up. What can we say about the eventual distribution of colored balls in the urn? Or about a series of trials in which we fill the urn and then start over again one hundred times?

- In each individual trial we have no idea what the eventual ratio of red to black balls will be; it could be 99.9% red, or 0.01% red, or anything in between. If we were to run 100 trials, we would probably get 100 different outcomes.
- In any particular trial, the ratio will eventually reach an equilibrium. Later draws in a series contribute only minutely to the distribution of balls in the urn. Thus, the distribution settles down onto a stable outcome.
- Sequence is thus crucial. Early draws in each trial, which have a considerable random element, have a powerful effect on which of the possible equilibria will actually emerge.

Mathematicians call this a Polya urn process. Its characteristic qualities stem from the fact that an element of chance (or accident) is combined with a decision rule that links current probabilities to the outcomes of preceding (partly random) sequences. Poly urn processes exhibit increasing returns or positive feedback. Each step along a particular path produces consequences which make that path more attractive for the next round. As such effects begin to accumulate, they generate a powerful virtuous (or vicious) cycle of self-reinforcing activity.

Increasing returns processes have quite intriguing characteristics, which Arthur (1994, 112–3) has summarized as follows.

1. **Unpredictability.** Because early events have a large effect and are partly random, many outcomes may be possible. We cannot predict ahead of time which of these possible end-states will be reached.

2. **Inflexibility.** The farther into the process we are, the harder it becomes to shift from one path to another. In applications to technology, a given subsidy to a particular technique will be more likely to shift the ultimate outcome if it occurs early rather than late. Sufficient movement down a particular path may eventually lock in one solution.4

3. **Nonergodicity.** Accidental events early in a sequence do not cancel out. They cannot be treated (which is to say, ignored) as “noise,” because they feed back into future choices. Small events are remembered.

4. **Potential path inefficiency.** In the long-run, the outcome that becomes locked in may generate lower pay-offs than a forgone alternative would have.

To this one can add a general point of particular interest to social scientists: These are processes in which sequencing is critical. Earlier events matter much more than later ones, and hence different sequences may produce different outcomes. In these processes, history matters.

If these characteristics are common in politics, then they carry major implications, both for the kinds of questions we should ask and the kinds of answers we should expect to find. Most important, they suggest the need to focus on the temporal dimensions of social processes. In searching for explanation, we need to think about causes and effects that are often separated in time, rather than focus exclusively on synchronic explanations (Harsanyi 1960; Stinchcombe 1968). Exploring these implications requires a review of recent work on increasing returns in economics.

### INCREASING RETURNS ARGUMENTS IN ECONOMICS

Economics has traditionally focused on the search for unique equilibria. The goal is attractive because it suggests a world of potential predictability and efficiency. Given knowledge of existing factor endowments and preferences, equilibrium analysis will point to a single optimal outcome. Moreover, because economists assume a context of decreasing marginal returns, this goal is potentially achievable. With decreasing returns, economic actions will engender negative feed-

---

2 The following discussion relies heavily on Arthur 1994, a collection of his ground-breaking essays on increasing returns and path dependence.

3 This case depicts a specific type of increasing returns process, in which the probability of a particular draw precisely equals the ratio between the two alternatives in the existing population. Arthur (1994) shows that many of the features of this case have a greater range of application, but not all of them. It is easy to model path dependent processes with only two equilibria (e.g., Hill 1997), which probably comes closer to capturing the essence of many path dependent processes in the social world.

4 This emerging stability represents a critical distinction between increasing returns processes and chaotic processes, which may generate no equilibrium. For an interesting discussion of this quite different framework, with applications to politics, see Fearon 1996.

5 Some social scientists, such as Stephen J. Gould, also have been drawn to arguments about path dependence, critical junctures, and punctuated equilibria in evolutionary biology (see especially Krasner 1989; Spruyt 1994). Without denying the relevance of this literature, I find it a less useful point of departure than the economists' focus on increasing returns. Most aspects of politics lack anything like the mechanism of natural selection that drives Darwinian theory (international relations and certain characteristics of electoral systems constitute important exceptions). Furthermore, socially created constructs of norms and formal institutions have no real analog in evolutionary theory. These constructs, however, are crucial features of politics and, as we shall see, are a critical element in social processes subject to increasing returns.
back, which will lead to a predictable equilibrium. A sharp rise in oil prices prompts increased conservation, exploration, and exploitation of other sources of energy, which will lead to a fall in oil prices. Each step away from equilibrium is more difficult than the one before. As Arthur (1994, 1) summarizes, negative “feedback tends to stabilize the economy because any major changes will be offset by the very reactions they generate. . . . The equilibrium marks the ‘best’ outcome possible under the circumstances: the most efficient use and allocation of resources.”

During the past fifteen years, however, this decreasing returns tradition has faced a mounting challenge. Economists have exhibited a growing interest in the idea of increasing returns. On a wide range of subjects—including the spatial location of production, the development of international trade, the causes of economic growth, and the emergence of new technologies—path dependence arguments have become prevalent. The ideas developed in this research are not entirely new, but they have been embraced and developed by prominent mainstream economists. This work has received considerable attention in leading journals. Douglass North, who places great emphasis on such arguments in his analysis of the development of modern capitalism, was awarded the Nobel Prize for economics.

The study of technology has provided the most fertile ground for arguments based on increasing returns. As Arthur (1994) and David (1985) have stressed, under conditions often present in complex, knowledge-intensive sectors, a particular technology may achieve a decisive advantage over competitors, although it is not necessarily the most efficient alternative in the long run. This occurs because each technology generates higher payoffs for each user as it becomes more prevalent. When a new technology is subject to increasing returns, being the fastest out of the gate (if only for reasons of historical accident) becomes critical. With increasing returns, actors have strong incentives to focus on a single alternative and to continue down a specific path once initial steps are taken in that direction. Once an initial advantage is gained, positive feedback effects may lock in this technology, and competitors are excluded. Path dependence arguments have been applied to the development of the “QWERTY” typewriter keyboard, the triumph of the light-water nuclear reactor in the United States, the battles between Betamax and VHS video recorders and between DOS-based and Macintosh computers, early automobile designs, and competing standards for electric current.  

Not all technologies, however, are prone to increasing returns. Arthur (1994) addresses not only the characteristics of such processes but also the conditions that give rise to them. Understanding these conditions is essential, as we shall see, because analytically similar circumstances occur frequently in the world of politics. Arthur’s characteristics provide a foundation for developing hypotheses about when increasing returns processes are likely to operate in the social world.

Arthur (1994, 112) argues that four features of a technology and its social context generate increasing returns.

1. **Large set-up or fixed costs.** These create a high pay-off for further investments in a given technology. With large production runs, fixed costs can be spread over more output, which will lead to lower unit costs. When set-up or fixed costs are high, individuals and organizations have a strong incentive to identify and stick with a single option.

2. **Learning effects.** Knowledge gained in the operation of complex systems also leads to higher returns from continuing use. With repetition, individuals learn how to use products more effectively, and their experiences are likely to spur further innovations in the product or in related activities.

3. **Coordination effects.** These occur when the benefits an individual receives from a particular activity increase as others adopt the same option. If technologies embody positive network externalities, then a given technology will become more attractive as more people use it. Coordination effects are especially significant when a technology has to be compatible with a linked infrastructure (e.g., software with hardware; automobiles with an infrastructure of roads, repair facilities, and fueling stations). Increased use of a technology encourages investments in the linked infrastructure, which in turn attracts still more users to the technology.

4. **Adaptive expectations.** If options that fail to win broad acceptance will have drawbacks later on, then individuals may feel a need to “pick the right horse.” Although the dynamic here is related to coordination effects, it derives from the self-fulfilling character of expectations. Projections about future aggregate use patterns lead individuals to adapt their actions in ways that help make those expectations come true.

This discussion of technology is important primarily because it clarifies a set of relationships characteristic of many social interactions. New social initiatives—such as the creation of organizations or institutions—usually entail considerable start-up costs; individuals, as well as organizations, learn by doing; the benefits of our individual activities or those of an organization are often enhanced if they are coordinated or “fit” with the activities of other actors or organizations; it is frequently important to bet on the right horse, so we adapt our actions in light of our expectations about the actions of others.

Although arguments about technology are probably the best known, economists have applied similar analyses of increasing returns processes in a striking range of economic contexts. Both Krugman (1991) and Arthur (1994, 49–67) point to the role of increasing returns in the spatial location of production. Given the

---

6 Many of these examples have been contested by critics who deny the empirical claim that superior technologies lost out. Since these criticisms raise broader issues about the usefulness of increasing returns arguments, I will postpone discussion until the end of this section.
importance of physical proximity in many aspects of economic life, agglomeration effects are widespread. That is, initial centers of economic activity may act like a magnet and influence the locational decisions and investments of other economic actors. Established firms attract suppliers, skilled labor, specialized financial and legal services, and appropriate physical infrastructure. The concentration of these factors may in turn make the particular location attractive to other firms that produce similar goods. So do social networks, which allow for easy exchange of information and expertise. Increasing returns arguments help explain the prevalence of pockets of specialized economic activity, from Silicon Valley to the high-end textile manufacturers of northern Italy. Krugman (1991, 80) concludes: “If there is one single area of economics in which path dependence is unmistakable, it is in economic geography—the location of production in space. The long shadow cast by history over location is apparent at all scales, from the smallest to the largest—from the cluster of costume jewelry firms in Providence to the concentration of 60 million people in the Northeast Corridor.”

These claims closely parallel recent analyses of international trade, an area in which arguments about increasing returns have gained wide acceptance. Researchers began by focusing on economic trends that appeared anomalous from the perspective of traditional trade theory—most notably, the explosion of intraindustry trade after World War II (Krugman 1996). If comparative advantage results from “natural” features of different countries, then one would expect most trade to occur between quite different countries, such as North-South trade of manufactured goods for raw materials. Yet, most trade is North-North, including extensive exchanges within particular industries. This pattern suggests a puzzle: Why have broadly similar countries developed highly specialized niche comparative advantages?

Increasing returns provide an answer. Knowledge-intensive sectors are prone to positive feedback. Countries that gain a lead in a particular field, for whatever reason, are likely to consolidate that lead over time. The result is a high degree of specialization. Even countries with similar initial endowments develop divergent areas of economic strength. Comparative advantage is not simply given, it is often created through a sequence of events over time.

It is worth noting that this research on trade has been used to derive some controversial policy implications. If first-mover advantages are significant, then free trade may not be an optimal policy for a country whose trade partners are willing to subsidize emerging sectors. Under certain (restricted) conditions, a policy of picking winners may make economic sense (Krugman 1996; Tyson 1993). There remains considerable dispute about the significance of such opportunities for strategic intervention. Krugman, for instance, maintains that they will appear relatively infrequently, not so much because path dependence is rare, but because governments will not be able to identify winners ex ante. Whatever the appropriate policy implications may be, however, the relevance of increasing returns processes to the economics of trade is now widely accepted.7

Economists also have applied increasing returns arguments to economic change more broadly. The most prominent development in recent discussions of economic growth centers on “endogenous growth” theory (Romer 1986, 1990). Economists in the 1980s became puzzled by growth rates (notably in developed countries after World War II) far greater than could be explained by measured increases in inputs of capital and labor. Romer and others argue that increasing returns associated with economic applications of knowledge help account for the anomaly. Unlike capital and labor, many aspects of knowledge are nonrival—their use in one firm does not prevent their use in another. A single gain in knowledge can be applied in many settings and can lead to dramatic improvements in productivity. Economic growth generates the positive feedback that defines increasing returns processes. A somewhat different analysis of growth based on increasing returns emphasizes the importance of complementarities (Milgrom and Roberts 1990). Various economic activities (e.g., in information technology) are complementary to other related activities. Improvements in a core activity can spill over by improving related parts of the economy (lowering costs or increasing productivity). These improvements in turn may increase the attractiveness of the core activity.

Economists are now applying increasing returns arguments to a wide range of important economic phenomena, but Douglass North’s application to issues of institutional emergence and change is perhaps most important for students of politics. North (1990a, 95) argues that all the features identified by Arthur in investigations of increasing returns in technology can be applied to institutions. In contexts of complex social interdependence, new institutions often entail high fixed or start-up costs, and they involve considerable learning effects, coordination effects, and adaptive expectations. Established institutions generate powerful inducements that reinforce their own stability and further development.

North emphasizes that not just single institutions are subject to increasing returns. Institutional arrangements induce complementary organizational forms, which in turn may generate new complementary institutions. For social scientists interested in paths of development, the key issue is often what North calls “the interdependent web of an institutional matrix.” This matrix, he emphasizes, “produces massive increasing returns” (North 1990a, 95). Path dependent processes will often be most powerful not at the level of individual organizations or institutions but at a more macro level that involves complementary configurations of organizations and institutions (Hall and Soskice 2000; Katznelson 1997).

7 As Krugman (1996, 109–10) notes, in the American Economic Association’s classification system for journal articles, one now finds “models of trade with increasing returns and imperfect competition” alongside the category for “conventional trade models.”
This argument provides the core to North’s sweeping reinterpretation of economic history. The central puzzle motivating North’s inquiry is the limited convergence of economic performance across countries over time. Neoclassical theory suggests that laggard countries should readily adopt the practices of high performers, which would induce fairly rapid convergence, but this does not happen. According to North, path dependent development of institutional matrices explains the anomaly of continued divergence in economic performance. Once in place, institutions are hard to change, and they have a tremendous effect on the possibilities for generating sustained economic growth. Individuals and organizations adapt to existing institutions. If the institutional matrix creates incentives for piracy, North observes, then people will invest in becoming good pirates. When institutions fail to provide incentives to be economically productive, there is unlikely to be much economic growth.

For political scientists, North’s insight is crucial for two reasons. First, he highlights the parallels between characteristics of technology and certain characteristics of social interactions. In this context, it is worth noting that Arthur’s arguments about technology are not really about the technology itself but about the characteristics of a technology in interaction with certain qualities of related social activity. Second, North rightly emphasizes that institutional development is subject to increasing returns. Indeed, it is the role of path dependence in explaining patterns of institutional emergence, persistence, and change that may be of greatest significance for political scientists.

The dialogue surrounding increasing returns in economics is the impassioned discourse of an emerging paradigm. Economists talk of “new” growth theory, “new” trade theory, and so on—all based on arguments involving increasing returns. Yet, despite the prevalence of such arguments and the intellectual excitement associated with them, there are excellent reasons to believe that the range of application should be at least as wide in politics as in economics. To understand why, it is helpful to consider the major objections to increasing returns arguments that have recently surfaced in economics. The discussion will clarify the sources of path dependence and identify social mechanisms that might offset such processes. This clarification provides a useful analytical bridge to an investigation of path dependent processes in politics.

In a forceful critique, Liebowitz and Margolis (1995) raise some tough questions about the literature on increasing returns. Two aspects of their argument are relevant here. They emphasize that only “remediable” path dependence is really of theoretical significance, and market mechanisms ensure that remediable path dependence is rare. I will take up each argument in turn. Following Williamson (1993), Liebowitz and Margolis (1995) distinguish remediable and nonremediable path dependence. The latter occurs if there are no feasible improvements in the path, either now or in the past. Nonremediable path dependence “stipulates that intertemporal effects propagate error” (p. 207). With hindsight, we wish that some other alternative had been chosen. Yet, Liebowitz and Margolis question whether this type of path dependence has profound implications. If we acted as best we could with the information available at the time, then the mistake was unavoidable, and we cannot reasonably describe the outcome as inefficient. Liebowitz and Margolis argue that the only kind of path dependence with major ramifications is path dependence that is potentially remediable.

Is their dismissal of nonremediable path dependence convincing? As Williamson (1993) notes, for policy purposes remediability is likely to be an appropriate standard. Recognizing the existence of path dependence may not help policymakers much if they do not know how to identify it ex ante.9 But this objection loses its force if our purpose is instead to understand—perhaps ex post—why aspects of societies move in particular directions and the consequences of such movements. And, of course, it is precisely these questions about causality that are the central preoccupation of most social scientists.

The second part of the Liebowitz and Margolis analysis is the claim that remediable path dependence is rare. Their argument is straightforward. If one of two options is superior in the long run but not in the short run, then market arrangements will generally assure the adoption of the superior path. The ability of private actors to capture the returns from long-term investments prevents bad choices. Institutions of property rights, provisions for patents, and extensive capital markets mean that options with low short-run pay-offs will nonetheless receive the support they deserve. Economic actors, in short, calculate in the shadow of the future and are, thus unlikely to indulge in myopic, short-term maximizing behavior at their own long-term expense.

This argument has considerable merit,10 but how much merit depends on the strength of these mechanisms for overcoming short-term thinking or free-riding. Although Liebowitz and Margolis are more than a little complacent about the capacity of these various market mechanisms, it is wise to leave those issues to economists. Two objections, however, are critically important. First, arguments about the farsightedness of markets seem to apply to only some types of path dependence in the economy. The Liebowitz and Margolis critique focuses on the decisions of firms to invest in particular technologies or products. In most of the illustrations discussed earlier (e.g., spatial agglomerations, trade specialization, endoge-

---

9 As noted before, precisely for this reason Krugman and others question those who make broad claims about the implications of increasing returns arguments for trade policy.

10 Indeed, Arthur (1994, 28, fn 11) explicitly recognizes this possibility, although as far as I know he does not systematically pursue the implications.
nous growth), however, many of the benefits of increasing returns are external to individual firms and cannot be fully captured by individual investors and entrepreneurs. Thus, the mechanisms identified by Liebowitz and Margolis are unlikely to ensure that the best long-term outcome will be selected.

Perhaps more important, the Liebowitz and Margolis (1995) argument has little relevance to the development of institutions, which are also subject to increasing returns. Private actors cannot obtain patents or venture capital to capture the long-term economic gains from constructing key economic institutions. Indeed, the Liebowitz and Margolis argument simply assumes the presence of institutions that support market mechanisms. Also, their argument does not seem to have much relevance for North’s argument about the presence in particular polities of networks or matrices of institutions and organizations. The fact that they do not even cite North’s work is telling. North maintains that path dependent processes of institutional development are crucial to the evolution of particular market economies. Far-sighted financial markets are of limited help in triggering such institutional development; to a large extent, they are its product.

The failure of Leibowitz and Margolis to address issues of institutional development in economics points to a more fundamental objection. Even if one accepts their analysis regarding the economic sphere, their arguments still have limited relevance for political scientists. However strong market mechanisms for “far-sightedness” may be, they are almost certainly far weaker in politics. I explain why in the next section.

MOVING FROM ECONOMICS TO POLITICS: THE APPLICABILITY OF INCREASING RETURNS ARGUMENTS

Microeconomic theory illuminates important features of the political landscape in fields ranging from the study of party competition, to the formation of interest groups and social movements, to voting and legislative behavior. The value of economists’ theoretical exports is greatly enhanced, however, if the political science importers take careful account of the distinctive features of the “local” environment. As Terry Moe (1990, 119) states in a related context: “The real problem is to try to identify those essential features of politics that might serve as a foundation for theory, a foundation that can take advantage of the new economics without being overwhelmed or misdirected by it.” Arguments drawn from economics must be sensitive to the quite different nature of the political world (Lindblom 1977; Moe 1984, 1990; North 1990b).

Politics differs from economics in many ways. The key is to specify which aspects are most relevant to an investigation of the sources and consequences of path dependence. Following a brief summary of the distinctive tasks of the political arena, this discussion is divided into two parts. The first considers four prominent and interconnected aspects of politics that make this realm of social life conducive to increasing returns processes: (1) the central role of collective action; (2) the high density of institutions; (3) the possibilities for using political authority to enhance asymmetries of power; and (4) its intrinsic complexity and opacity. After briefly explicating each, I will discuss their relevance to path dependence. Each of these features makes increasing returns processes prevalent in politics.

Second, I explain why the ameliorative mechanisms that Liebowitz and Margolis identify in economic systems are often ineffective in offsetting path dependence in politics. Three characteristics of politics change the picture considerably: the absence or weakness of efficiency-enhancing mechanisms of competition and learning; the shorter time horizons of political actors; and the strong status quo bias generally built into political institutions. Each of these features makes increasing returns processes in politics particularly intense. They increase the difficulty of reversing the course down which actors have started. Increasing returns processes are now central to economic theory and the argument here is that these dynamics will be very widespread and often more difficult to reverse in politics.

For my purposes, the fundamental feature of politics is its preoccupation with the provision of public goods. Such goods are distinguished by jointness of supply (the production costs for the good are unaffected or only modestly affected by the number of those consuming it) and nonexcludability (it is very costly or impossible to limit consumption to those who have paid for a good). These features, which are extremely widespread in modern life, make public goods—from national defense to environmental protection—difficult to provide through markets. Nonexcludability creates free-riding incentives, since individuals will receive the benefits of a public good whether or not they contribute to its production. Jointness of supply means that private markets will underproduce the good in question, since private actors tend to consider only the benefits to themselves.

These characteristics of public goods help explain why the central features of political systems are compulsory rather than voluntary. The exercise of authority, generally combined with a complex array of complementary institutions designed to circumscribe and legitimate that authority, is necessary to generate collective provision. Legally binding rules are not just a foundation for political activity (like property rights in the economy). They are instead the very essence of politics (Lindblom 1977; Moe 1990). This key quality of politics has a number of repercussions for the character of political life, each of which is relevant for an assessment of tendencies toward path dependence.

The Collective Nature of Politics

A quick contrast with economic markets can highlight the prevalence of collective action in politics. Suppose

---

11 In most cases, the goods in question are not "pure" public goods—a fact that complicates the analysis but does not alter my basic claims. For discussions, see Mueller 1989, chap. 2, and Corneo and Sandler 1996.
you work for a firm with an annoying boss and bad pay. You have a clear option: Seek work elsewhere, either at one of a large number of other firms or by setting up business on your own. Your ability to move depends on the state of the labor market, but the existence of competitive options sets clear limits on how annoying your boss can afford to be and on how bad the pay can get.

Or suppose you invent a great new product. Assuming that you can get financial backing (which you should be able to do—it is a great idea, and the market generates a ready supply of venture capitalists), your prospects are good. Nothing stops you from going into business or selling the idea to someone. Either way, the new, superior product gets to see the light of day, and you reap considerable benefits from your innovation.

The setting of consumers, at least in the textbook case, is similarly atomistic. My decisions are essentially independent of my expectations regarding the choices of other consumers. There is no need for explicit attempts to coordinate behavior; the market simply aggregates the isolated decisions of individuals.

These highly stylized examples illustrate the flexibility, fluidity, and atomization of economic markets. In contrast, political "markets" are generally far from flexible and fluid. In politics, the consequences of my actions are highly dependent upon the actions of others. What I get depends not just on what I do, but (mostly) on what others do. Following Olson's (1965) path-breaking work, students of politics have long recognized the "logic of collective action." Most of the "goods" produced in politics are public goods; it is difficult to limit their consumption to those who helped provide them. As a result, individuals will have a strong tendency to free-ride. Creating conditions favorable to collective action is a principal issue in political life.

The problem is not limited to the fact that the public sector produces public goods. Given the reliance of politics on mechanisms of collective decision backed by authority, laws themselves have the character of public goods for those who benefit from them. In the words of Marwell and Oliver (1993, 42), "influencing government policy almost always has very high jointness of supply." These circumstances generate major collective action problems.

There is another reason political action frequently requires coordination. Many of the goals pursued by political actors have a "lumpy" or "winner-take-all" quality (politicians seeking reelection, coup plotters, and lobbyists either win or lose; legislation either passes or is rejected). Unlike economic markets, in which there usually is room for many firms, in politics finishing second may not count for much. Indeed—the Mensheviks in 1917 come to mind—it can be extremely problematic. Again, the effectiveness of my actions depends heavily on the actions of others. This is less true of some aspects of politics—such as answering an opinion poll question or voting—than others. Even in voting, however, the lumpiness of election outcomes (in the absence of a pure system of proportional representation) means that the actions of a person who does not want to "waste" her vote may well turn on what she expects others to do.

A crucial feature of most collective action in politics is the absence of a linear relationship between effort and effect. Instead, collective action frequently involves many of the qualities conducive to positive feedback (Marwell and Oliver 1993). A central reason is the prevalence of adaptive expectations. When picking the wrong horse may have very high costs, actors must constantly adjust their behavior in the light of how they expect others to act. Whether you put energy into developing a new party, or join a potential coalition, or provide resources to an interest group may depend to a considerable degree on your confidence that a large number of other people will do the same. In addition, many types of collective action involve high start-up costs, which reflects the fact that considerable resources (material or cultural) need to be expended on organizing before the group becomes self-financing.

That collective action processes in politics are very often subject to increasing returns explains why social scientists are often struck by the considerable stability of patterns of political mobilization over time. Lipset and Rokkan's (1967) work on political parties in Europe exemplifies this dynamic: Key historical junctures produced major political cleavages. These political divisions became organized into political parties. Once they have surmounted initial start-up costs and fueled processes of adaptive expectations, these parties are reproduced through time, which generates "frozen" party systems.

Recent work by Skocpol (1999) on extensive voluntary associations in the United States provides additional strong evidence of the organizational persistence that can result from positive feedback. Skocpol identified all such organizations that had ever enrolled more than 1% of the American population (or half that amount for single-gender groups) and tracked them over time. The results, which cover 58 groups since the 1830s, reveal striking organizational continuities. Although some crossed the 1% threshold only for a relatively short period, 26 remain above it today. Of these, 16 had reached the 1% mark by the 1940s, and a number of them stretch back much farther. A large number have fallen from the list, but they managed to stay on it for many decades. Among the 40 organizations founded before 1900, 19 stayed above the 1% mark for at least five decades. Ten of the 40 are still above that threshold, a century or more after their founding. In short, despite massive social, economic,

---

12 This represents a critical difference between economics and politics, but a number of important qualifications need to be made. For instance, consumption often involves significant externalities, which make consumer choices interdependent. As already noted, conditions of independent consumption often do not apply to high-tech products, which frequently involve network externalities. For a good discussion of some of these complications, see Hirsch 1977.

13 It should be emphasized that this evidence understates organizational persistence, since many groups that fall below the demanding 1% threshold nonetheless continue to have a very large membership. Also, they may have existed as quite large organizations for long periods before initially crossing the threshold.
and political changes over time, self-reinforcing dynamics associated with collective action processes mean that organizations have a strong tendency to persist once they are institutionalized.

The Institutional Density of Politics

As much recent work in political science stresses, efforts to coordinate actors in the pursuit of public goods often require the construction of formal institutions. Once established, these institutional constraints can apply to all—those who do not approve as well as those who do—and they are backed up, ultimately, by force. The exit option, so central to the workings of the market, is often unavailable (or prohibitively costly) to actors who feel poorly served by existing political arrangements. In politics, institutional constraints are ubiquitous. Politics involves struggles over the authority to establish, enforce, and change the rules governing social action in a particular territory. In short, much of politics is based on authority rather than exchange. Both formal institutions (such as constitutional arrangements) and public policies place extensive, legally binding constraints on behavior.

Although unorthodox, the inclusion of public policies as well as formal institutions in this formulation is important (Pierson 1993). Policies are generally more easily altered than the constitutive rules of formal institutions, but they are nevertheless extremely prominent constraining features of the political environment. Policies, grounded in law and backed by the coercive power of the state, signal to actors what has to be done and what cannot be done, and they establish many of the rewards and penalties associated with particular activities. Most policies are remarkably durable (Rose 1990). Especially in modern societies, extensive policy arrangements fundamentally shape the incentives and resources of political actors.

That such institutions are prone to increasing returns is implicit in much recent research on institutions. Scholars emphasize how institutions can help actors overcome various dilemmas arising from collective choice situations—especially the need to coordinate their behavior by disciplining expectations about the behavior of others. What is absent or downplayed, however, is a recognition that these characteristics render processes of institutional development path dependent.

As already discussed, North highlights how institutions induce self-reinforcing processes that make reversals of course increasingly unattractive over time. In contexts of complex social interdependence, new institutions and policies are costly to create and often generate learning effects, coordination effects, and adaptive expectations. Institutions and policies may encourage individuals and organizations to invest in specialized skills, deepen relationships with other individuals and organizations, and develop particular political and social identities. These activities increase the attractiveness of existing institutional arrangements relative to hypothetical alternatives. As social actors make commitments based on existing institutions and policies, their cost of exit from established arrangements generally rises dramatically.

Political Authority and Power Asymmetries

In the famous community power debate of the 1960s and 1970s, Bachrach and Baratz (1962) and Lukes (1974) argued persuasively that power asymmetries are often hidden from view; where power is most unequal, it often does not need to be employed openly. Pluralist critics essentially countered that a systematic evaluation of such claims was impossible (Polity 1963; Wolfinger 1971). Although he does not frame the issue quite this way, Gaventa (1980) demonstrates that power asymmetries can reflect the operation of positive feedback processes over substantial periods. Increasing returns processes can transform a situation of relatively balanced conflict, in which one set of actors must openly impose its preferences on another set (“the first face of power”), into one in which power relations become so uneven that anticipated reactions (“the second face of power”) and ideological manipulation (“the third face”) make open political conflict unnecessary. Thus, positive feedback over time simultaneously increases power asymmetries and renders power relations less visible.

The allocation of political authority to particular actors is a key source of this kind of positive feedback. Indeed, this represents a source of path dependence quite distinct from those discussed by Arthur and North. When certain actors are in a position to impose rules on others, the employment of power may be self-reinforcing (Mahoney 1999). Actors may use political authority to generate changes in the rules of the game (both formal institutions and various public policies) designed to enhance their power. Relatively small disparities in political resources among contending groups may widen dramatically over time as positive feedback sets in.

The Complexity and Opacity of Politics

Economic theory is built in large part around the useful and plausible assumption that actors seek to optimize and are relatively good at it. Firms operate to maximize profits. The metric for good performance is relatively simple and transparent. Prices send strong signals that facilitate the analysis of how various features of the economic environment affect firm performance. Observable, unambiguous, and often quantifiable indicators exist for many of these features. Workers can to mean previous outlays that cannot be recovered and should be regarded as irrelevant to current choices among options. The whole point of path dependence, however, is that these previous choices often are relevant to current action. In cases of increasing returns, social adaptations represent investments that yield continuing benefits. Actors may be locked into a current option because massive new investments may be required before some theoretically superior alternative generates a higher stream of benefits.
obtain fairly good information on the wages and working conditions on offer from different firms. Consumers, too, are reasonably adept at navigating most aspects of the economic world. Links between choices and outcomes are generally clear: Take a new job and your income rises; buy a car and your savings account balance shrinks. The quality of goods is usually evident in relatively short order, and repeated purchases allow consumers to sample alternatives.

Of course, one can add many complications to this simple picture of the economic realm. The market is often highly complex and confusing. Yet, the clarifying role of prices, the prevalence of repeated interactions, the absence of a need to coordinate many of one’s economic decisions with those of other actors, and the presence of relatively short causal chains between choices and results greatly facilitate the efforts of economic actors to correct mistakes over time.

Politics is a far, far murkier environment (Moe 1990; North 1990b). It lacks anything like the measuring rod of price. Political actors pursue a range of goals. Furthermore, it is often very hard to observe or measure important aspects of political performance. And, if we believe that a system is not performing well, it is still more difficult to determine which elements in these highly complex systems are responsible and what adjustments would lead to better results. The reliance on elaborate procedures to handle collective choice situations in politics is inescapable, but it undermines transparency, that is, it greatly increases transaction costs (Cornes and Sandler 1996; Mueller 1989). The complexity of the goals of politics as well as the loose and diffuse links between actions and outcomes render politics inherently ambiguous.

Even if mistakes or failures in politics are apparent, improvement through trial-and-error processes is far from automatic. Many participants in politics (voters, members of interest groups) engage in activities only sporadically. Their tools of action are often crude, such as the blunt instrument of the vote, and their actions have consequences only when aggregated. There may be long lags and complex causal chains connecting these political actions to political outcomes. The result is that mistaken understandings often do not get corrected.

The point is not that learning never occurs in politics. Rather, learning is very difficult and cannot be assumed to occur. Instead, understandings of the political world should themselves be seen as susceptible to path dependence. Drawing on work in both cognitive psychology and organizational theory, researchers argue that actors who operate in a social context of high complexity and opacity are heavily biased in the way they filter information into existing “mental maps” (Arthur 1994; Denzau and North 1994). Confirming information tends to be incorporated, and disconfirming information is filtered out. Social interpretations of complex environments like politics are subject to positive feedback. The development of basic social understandings involves high start-up costs and learning effects; they are frequently shared with other social actors in ways that create network effects and adaptive expectations. The need to employ mental maps induces increasing returns. This is true both at the individual level and at the group level, as “communities of discourse” often come to share and reproduce a similar ideology (Wuthnow 1989).15

This recent work converges with the long-standing views of those who study political culture as well as the recent contributions of cognitive science.16 Once established, basic outlooks on politics, ranging from ideologies to understandings of particular aspects of governments or orientations toward political groups or parties, are generally tenacious. They are path dependent.

There are, then, compelling reasons to believe that political life will often be marked by dynamics of increasing returns. Tendencies toward positive feedback characterize four processes central to political environments: collective action, institutional development, the exercise of authority, and social interpretation. In each case, there are reasons to anticipate that steps in a particular direction can trigger a self-reinforcing dynamic. This conclusion should be underlined. By itself, it suggests why increasing returns is a critical concept for those who seek to understand the sources of political stability and change. A recognition that self-reinforcing processes are significant is shaking up economics, and political scientists have at least as great a need to consider their implications.

There is also reason to believe that these effects in politics are often particularly intense. In the remainder of this section I consider why it is frequently more difficult to reverse course in politics than it would be in economics. Economists argue that the market provides two powerful mechanisms for exiting problematic paths: competition and learning. Competitive pressures in a market society mean that new organizations with more efficient structures will develop and eventually replace suboptimal organizations (Alchian 1950). Learning processes within firms also can lead to correction. According to Williamson (1993, 116–7), one can rely on

15 Wuthnow’s (1989) subtle analysis of the comparative development of ideologies, with its emphasis on relatively brief periods of historical openness followed by processes that select and then institutionalize a particular track of ideological development, is broadly consistent with the framework suggested here.

16 Consider the statement by Mannheim (1952, 298) in his famous essay on generations: “It is of considerable importance for the formation of consciousness which experiences happen to make those all-important ‘first impressions’. . . . Early impressions tend to coalesce into a natural view of the world.”

17 Indeed, as marketers know well, path dependent cognitive effects are evident even in the less ambiguous world of consumption. This is why advertisers covet the attention of youngsters, who have yet to make definitive (and resilient) choices. A telling recent example is the marketing effort of the National Football League, which is alarmed by indications that youngsters are increasingly drawn to basketball and soccer. A former MTV executive now working on special events speaks the language of increasing returns: “It’s all about getting a football . . . into a kid’s hands as soon as you can. Six years old, if possible. You want to get a football in their hands before someone puts a basketball in their hands, or a hockey stick or a tennis racquet or a golf club” (Seabrook 1997, 47).
the “far-sighted propensity” or “rational spirit” that economics ascribes to economic actors. . . . Once the unanticipated consequences are understood, these effects will thereafter be anticipated and the ramifications can be folded back into the organizational design. Unwanted costs will then be mitigated and unanticipated benefits will be enhanced. Better economic performance will ordinarily result.

Of course, neither mechanism represents a guaranteed corrective in the context of increasing returns. Options that gain a head start will often reinforce themselves over time, even if they have serious shortcomings. What I wish to stress, however, is that Williamson’s corrective mechanisms are even less effective when one shifts from firms in private markets to the world of political institutions (Moe 1984, 1990; Pierson n.d.a). This is clearest for mechanisms of competition. Political institutions rarely confront a dense environment of competing institutions that will instantly capitalize on inefficient performance, swooping in to carry off an institution’s “customers” and drive it into bankruptcy. Models of competition may be helpful for understanding some important aspects of politics (such as international relations and elections), but there can be little doubt that political environments are typically more “permissive” than economic ones (Krasner 1989).

As just discussed, the complexity and ambiguity of politics create serious problems for learning arguments. It may be appropriate to argue that politics sometimes involves learning processes, in which responses to public problems proceed in a trial-and-error fashion (Hall 1993; Heclo 1974). There is little reason, however, to think that this acts as a selection mechanism with anything like the efficiency-enhancing properties of market competition in economics or Darwinian natural selection in biology. Because political reality is so complex and the tasks of evaluating public performance and determining which options would be superior are so formidable, such self-correction is often limited.

Even when learning does occur, it faces additional hurdles. In Williamson’s (1993, 117) words, learning must still be “folded back into the organizational design.” All the barriers to change in systems subject to increasing returns become relevant: Long movement down a particular path will increase the costs of switching to some previously forgone alternative. Furthermore, in politics the pursuit of such change faces two additional obstacles: the short time horizons of political actors and the strong status quo bias associated with the decision rules that govern most political institutions. These factors often make path dependent effects particularly intense in politics.

**Time Horizons.** A statement attributed to David Stockman, budget director during the Reagan administration, is unusual among political decision makers only for its candor. Asked by an adviser in 1981 to address Social Security’s severe long-term financing problems, Stockman dismissed the idea out of hand. He explained that he had little interest in wasting “a lot of political capital on some other guy’s problem in [the year] 2010” (quoted in Greider 1982, 43).

Many of the implications of political decisions—especially complex policy interventions or major institutional reforms—only play out in the long run. Yet, political actors, especially politicians, are often most interested in the short-term consequences of their actions; long-term effects tend to be heavily discounted. The principal reason is the logic of electoral politics. Because the decisions of voters are taken in the short run, elected officials generally employ a high discount rate. They will pay attention to long-term consequences only when these become politically salient or when they have little reason to fear short-term electoral retribution. As John Maynard Keynes once noted, in the long run we are all dead; politicians in democratic politics have special reason to take that message to heart.

Political scientists have paid limited attention to the issue of time horizons. An interesting literature is developing on “credible commitments”—the attempt of political actors to create arrangements that facilitate cooperation by lengthening time horizons (North 1993; North and Weingast 1989; Shepsle 1991). We know relatively little about the time horizons of different political actors or about the institutional arrangements conducive to lowering their discount rates (i.e., increasing the political relevance of the future). Recent research suggests that particular institutional designs (such as independent central banks), which empower particular kinds of political actors (e.g., bankers), may succeed in lengthening time horizons in politics.

In general, however, such mechanisms are less effective in politics than in economics. As noted, the marketplace possesses some strong mechanisms for lengthening time horizons—especially property rights and capital markets. The mechanisms in politics are generally far weaker. Monitoring political behavior over time is difficult because indicators of performance are typically so limited. It is no accident that much of the generally optimistic rational choice discussion of “credible commitments” in politics has focused on relatively transparent economic issues (e.g., budget deficits, monetary policy). In these instances, performance indicators are clear, and behavior is easy to monitor. Although these issues are clearly important, it must be stressed that for reasons already noted they are fundamentally atypical of the kinds of matters dealt with in politics. Not only is monitoring often exceptionally difficult in politics, but also the relatively rapid turnover of key positions makes it hard to hold actors accountable. Politics, in short, lacks the characteristic property rights that facilitate the linkage of actors’ decisions over time in the economic sphere. In many cases, the long term is essentially beyond the political horizon. A statesman, Bismarck is said to have pronounced, is a politician who thinks about his grandchildren.

The different nature of time horizons in politics and in economics matters a lot. This can be seen by revisiting the Liebowitz and Margolis (1995) critique of path dependence. They properly point to key market institutions as a (partial) protection against (certain
kinds of) remediable path dependence. If it is believed that one option (say, Amazon.com) has greater long-term benefits, then investors should gravitate toward that option even if in the short term it will perform more poorly than an alternative. They argue that market mechanisms should allow the more efficient result.

In politics the outcome may well be different. Assume that the crucial decision maker is a politician up for reelection in two years. In this context, effects after the election cycle may not count for much. A politician who focuses on the short-term pay-off will choose a different option from that of far-sighted economic investors. This difference in time horizons has profound consequences. If time-horizons tend to be short, then we can expect that long-term costs and benefits will have a limited effect on the chosen path. Furthermore, once on a particular path, political actors will generally have powerful incentives to stay on it. Switching costs are typically borne in the short run, and the benefits will generally only accrue in the long run, that is, to someone else.

The Status Quo Bias of Political Institutions. Political arrangements are unusually hard to change. In the economic realm, an individual with a new idea for a product need only secure the financing to put it on the market. If enough consumers (choosing independently) find it sufficiently appealing, the product will be a success. Change can be engineered through competition against existing products. Similarly, those with property rights over a firm are generally in a strong position to remake their organizations as they choose. Lines of authority are clear, and the relevant decision makers are likely to share the same broad goal of maximizing profits.

By contrast, the key features of political life—public policies and (especially) formal institutions—are change-resistant. Both are generally designed to be difficult to overturn for two broad reasons. First, those who design institutions and policies may wish to bind their successors. According to Moe (1990), this reflects the problem of "political uncertainty." Unlike economic actors, political actors must anticipate that their political rivals may soon control the reins of government. To protect themselves, they may create rules that make preexisting arrangements hard to reverse. As Moe (1990, 125) puts it, designers "do not want 'their' agencies to fall under the control of opponents. And given the way public authority is allocated and exercised in a democracy, they often can only shut out their opponents by shutting themselves out too. In many cases, then, they purposely create structures that even they cannot control."

Second, in many cases, political actors also are compelled to bind themselves. The key insight of the "credible commitments" literature is that actors can often do better in the short run as well as the long run, if they remove certain options from their future menu. The economy of a country will grow faster, for instance, if a monarch can credibly commit himself to refrain from expropriating an excessive amount of the hard-earned wealth of his subjects (North and Weingast 1989). This can be done if he accedes to parliamentary control over the power to tax.

To constrain themselves and others, designers create large obstacles to institutional change. The barriers to reform may be extremely high, such as unanimity requirements in the European Union and multiple supermajorities to alter the U.S. Constitution. Of course, these obstacles facilitate forms of cooperation and exchange that would otherwise be impossible. The relevant point here is that this status quo bias characteristic of political systems reinforces the already considerable difficulties of moving off an established path. Combined with the lack of competitive mechanisms, the weakness of learning processes, and the short time horizons characteristic of politics, the bias means that increasing external tendencies in political development are often particularly intense.

Politics differs from economics in many ways. Applying tools of economic analysis to politics is treacherous unless these differences are systematically considered. In the case of arguments about path dependence, attention to the character of politics suggests a striking result. The political world is unusually prone to increasing returns. Both the prevalence and intensity of increasing returns processes suggest that path dependence arguments offer important insights for understanding political dynamics.

---

18 These long-term effects will count if an actor with longer time horizons (such as an interest group) can make them relevant to politicians, such as through campaign contributions or votes. The question is whether such mechanisms are anywhere near as effective as the capital markets operative in the economic sphere. In my view, there are strong reasons to be skeptical, but it is an issue that deserves considerable attention.

19 This assumes that the actors involved care about what the government does after their faction loses an election. For reasons just noted (and as Moe observes) the problem of long-term political uncertainty is likely to be of greater concern for interest groups than for politicians.

20 An important characteristic of political systems runs counter to this line of argument. Because politics is a powerful system for mobilizing coercive power, governments may at times be in a position to orchestrate a "jump" from one path to another. By employing sanctions, they can coordinate adjustments in a way that markets might never be able to achieve. For instance, the British government enacted a shift to the metric system that would have been difficult or impossible to engineer through the more atomistic mechanisms of the market. Governments are clearly capable on occasion of mobilizing resources for more dramatic changes in course. Such possibilities, however, should not be exaggerated. The metric example represents a modest instance of reversing path dependence. Switching costs were relatively low; the principal problem was one of coordination, of inducing everyone to make the switch at the same time. For this task, the authoritative rule-setting capacities of government are of great assistance. For reasons already discussed, it is much less evident that governments will generally be willing or able to engineer shifts to a different path when switching costs are high. Cases of fundamental or revolutionary reform in well-institutionalized political systems attract our attention precisely because they are so rare.
PATH DEPENDENCE AND THE STUDY OF POLITICS

To summarize briefly, in settings in which increasing returns or path dependent processes are at work, political life is likely to be marked by four features.

1. **Multiple equilibria.** Under a set of initial conditions conducive to increasing returns, a number of outcomes—perhaps a wide range—are generally possible.

2. **Contingency.** Relatively small events, if they occur at the right moment, can have large and enduring consequences.

3. **A critical role for timing and sequencing.** In increasing returns processes, when an event occurs may be crucial. Because earlier parts of a sequence matter much more than later parts, an event that happens “too late” may have no effect, although it might have been of great consequence if the timing had been different.

4. **Inertia.** Once an increasing returns process is established, positive feedback may lead to a single equilibrium. This equilibrium will in turn be resistant to change.

There are also good reasons to think that increasing returns processes are widespread in politics, since they will be characteristic in institutional development, collective action, the exercise of authority, and the emergence of our understandings of the political world.

If increasing returns processes are prevalent in politics, then there are fundamental theoretical implications. We need to change both the kinds of questions we ask about politics and the kinds of answers that we generate. With respect to questions, the most important implication is the need to focus on branching points and on the specific factors that reinforce the paths established at those points. Students of comparative politics and American political development have long been interested in critical moments. The dynamics of increasing returns lay out what Collier and Collier (1991, 31) term the “mechanisms of reproduction,” which carry and often amplify the effects of a critical juncture through time. Discussions of path dependence and critical junctures are often conflated, but as Hacker (1998) has emphasized, it is important to keep them distinct. Arguments about path dependence explain why particular historical junctures have lasting consequences. Yet, although claims about critical junctures seem to rest on assertions of increasing returns processes (otherwise it is not clear why the juncture is “critical”), the inverse does not hold. Path dependent arguments based on positive feedback suggest that not only “big” events have big consequences; little ones that happen at the right time can have major consequences as well.

A focus on increasing returns processes justifies a turn to history. At one level, of course, all social scientists agree that history matters. Current conditions, which influence current social outcomes, came into being in some way. Those earlier processes are relevant to a full understanding of contemporary social events. Yet, the standard assumption is that for most purposes we may safely put such issues aside. Looking back leads to the familiar problem of infinite regress. An exploration of each preceding event leads to the conclusion that some other previous occurrence was also part of the chain of necessary events, and so on. Social scientists need to break through the seamlessness of history somewhere, and the present is as good a place to do so as any. Homans (1967, 92–3) compares the situation to that faced by navy minesweepers, who need to know the magnetic charge of a ship. Such a charge results from an infinite range of small factors accumulated over the ship’s lifetime. For practical purposes, however, a simple expedient can be used: The current charge of the ship can be measured. If the task is to understand the ship’s vulnerability to mines, one can simply cut through the Gordian knot of historical regress.²¹

For many purposes, this is an appropriate approach. Social scientists often have good reason to focus on synchronic causality—to try to understand how variations in current variables affect present social outcomes. When increasing returns processes are significant, however, such a strategy may be problematic. Increasing returns arguments rest on a conception of “historical causes” (Harsanyi 1960; Ikenberry 1994; Stinchcombe 1968, 103–18), that is, some original ordering moment triggered particular patterns, and the activity is continuously reproduced even though the original event no longer occurs. Under conditions of path dependence it is true that current circumstances in some sense “cause” current outcomes, but a focus on these simultaneous occurrences is highly misleading. It provides a “snapshot” explanation for what should be seen as a moving picture. The necessary conditions for current outcomes occurred in the past. The crucial object of study becomes the critical juncture or triggering events, which set development along a particular path, and the mechanisms of reproduction of the current path—which at first glance might seem commonplace or at least analytically uninteresting.

An awareness of increasing returns processes can change not only the questions we ask but also the answers we provide. Put differently, an understanding of increasing returns can be a fruitful source of hypotheses about the sources of social outcomes. One merit of increasing returns arguments is that they provide a plausible counter to functionalist explanations in political science, which often go unchallenged. Although not always explicitly stated, functionalist arguments are prevalent among political scientists. They are common, for instance, among those who emphasize the rational choices of individual actors that underlie political activity and the reasonably efficient nature of collective responses to social needs (Keohane 1984; Shepsle 1986; Weingast and Marshall 1988).

Functionalist arguments take the following form: Outcome X (e.g., an institution, policy, or organization) exists because it serves the function Y. In a world of purposive actors, it may indeed be the case that the

²¹ For a discussion of Homans’s argument, see Knapp 1984, 43–5.
effects of an institution have something to do with an explanation for its emergence and persistence. Arguments about increasing returns, however, suggest the large dangers in any assumption that an institution arose because it serves some particularly useful purpose. Thinking in functionalist terms about an institution, policy, or social organization may be a good way to derive causal hypotheses, but functional accounts are far from being the only plausible ones. Many alternatives to the outcome in question might have been possible, and a dynamic of increasing returns may have locked in a particular option even though it originated by accident, or the factors that gave it an original advantage may have long since passed away. Rather than assume relative efficiency as an explanation, we have to go back and look. Thus, recognizing the possibility of path dependence necessarily draws social scientists to an investigation of history, if only to evaluate the validity of functionalist assertions.

More positively, an investigation of path dependence can provide a basis for developing important hypotheses about the sources of political stability and change. To repeat, Arthur's work on increasing returns is ground-breaking not simply because it describes the characteristics of these processes; it identifies conditions conducive to path dependence. The major ambition of this essay, building on North's work, is to begin the process of adapting these arguments to the study of politics. Doing so requires careful attention to the distinctive features of the political world, such as its intrinsic ambiguity, the prevalence of change-resistant institutions, the prominence of collective action problems, and the prospects for using political authority to amplify asymmetries of power. Not all aspects of political life are subject to increasing returns. Furthermore, this article has highlighted the more specific features of political environments that are likely to influence the initiation and reinforcement of increasing returns processes. In short, this is fertile territory for developing new propositions about the conditions that facilitate or impede various types of political change.

Consider one example. A prominent theme in recent research in comparative political economy is "varieties of capitalism." Despite increasing international economic interdependence, which seems to generate pressures toward convergence, the advanced industrial societies continue to exhibit fundamental differences in their core institutional structures (Berger and Dore 1996; Hall 1999; Hollingsworth and Boyer 1997; Soskice 1999). To date, this literature has done a better job of identifying and describing the diversity than it has of explaining what generates and sustains it. Hall and Soskice (2000) have made an important step forward by emphasizing the role of institutional complementarities. The benefits of particular economic institutions and organizations are increased if they operate in an environment populated by specific kinds of institutions and organizations.

The "varieties of capitalism" analysis persuasively illuminates distinct equilibria in different economies, but it does not address how these distinct equilibria emerge. From the current analysis, one can easily see why the elaborate production systems of modern economies would be subject to increasing returns. Start-up costs, not just for new firms but more fundamentally for the key organizations and institutions that link private actors, are enormous. Organizations, and the formal and informal arrangements (both public and private) that help structure their interactions, create densely linked institutional matrices. Economic and social organizations and political institutions (both basic constitutional arrangements and public policy frameworks) have coevolved over extended periods. Coordination effects are widespread; particular courses of action make sense because of anticipated actions of others in the system. Firms have developed sophisticated strategies suitable to the particular institutional matrix they confront, that is, tremendous amounts of learning by doing have occurred over time in these complex systems. In short, national economic systems are highly path dependent. They are likely to exhibit substantial resilience, even in the context of major exogenous shocks, such as recent changes in the global economy.

In addition to highlighting particular causal processes that generate or sustain positive feedback, increasing returns arguments also direct attention to hypotheses explicitly based on timing and sequence. Under conditions conducive to path dependence, the same event (e.g., an exogenous shock such as depression or war) may have a different effect depending on when in a sequence of events it occurs (Collier and Collier 1991; Ertman 1996). Skowronek (1993) persuasively argues that we cannot understand the opportunities, constraints, and demands a president faces without placing him within a sequence of presidencies that support or oppose the dominant coalition of a particular period. Path dependence arguments provide a stronger foundation for Tilly's (1984, 14) claim that "when things happen in a sequence affects how they happen."

This highlights the broader theoretical significance of path dependence arguments: They can help political scientists think more clearly and explicitly about the role of time, and history, in social analysis. This is crucial because some claim to be witnessing a "historic turn" in the social sciences (McDonald 1996), but there is much confusion about what such a turn might mean. For some, particular historical outcomes are of intrinsic interest. For many, historical analysis is essentially a method, a way to expand the universe of cases that one can use to illustrate supposedly general theoretical models. The claim here is quite different. We should turn to history because important aspects of social reality can best be comprehended as temporal processes. It is not the past per se but the unfolding of processes over time that is theoretically central.

The main properties of increasing returns processes provide considerable support for many of the key claims of "historical institutionalist" analyses in political science. The phrase is a fortunate one, as it captures two critical themes explored here. This work is historical because it recognizes that political development must be understood as a process that unfolds over time.
It is institutionalist because it stresses that many of the contemporary political implications of these temporal processes are embedded in institutions—whether formal rules, policy structures, or norms.

Of these two elements, the institutional side generally has received greatest attention. Despite notable exceptions (Katznelson 1997; Orren and Skowronek 1994; Skocpol 1992; Skowronek 1993; Thelen 1999), the significance of temporal processes in historical institutionalist analysis is often left implicit or downplayed. Nevertheless, empirical work in this tradition highlights the need to examine temporal processes in order to explain important political outcomes. Historical institutionalist scholarship often emphasizes critical moments in politics, distinctive developmental sequences, and the rigidities that make it difficult for social actors to escape from established paths.

Of course, recent works of historical institutionalism build on a tradition of attention to history in the social sciences. Particularly for those who want answers to critical questions that grow out of the experiences of real polities, the turn to history has been common. Issues of timing, sequence, and critical junctures figure prominently in this body of work. Among many such studies, Gerschenkron’s (1962) study of industrialization and state-building and Lipset and Rokkan’s (1967) analysis of the formation of party systems are two classic examples. Indeed, it is fair to ask whether incorporating the concepts of increasing returns and path dependence into the study of politics is akin to the man who discovered that he had been speaking prose all his life. Is path dependence merely a trendy name for old ideas?

Discussions of path dependence would be worthwhile even if they did no more than focus the attention of afad-prone discipline on the insights and continuing relevance of this earlier body of work. Yet, there is every reason to believe that the concept can do more. Knowledge of the dynamics of increasing returns processes can greatly sharpen our understanding of why particular junctures (and which aspects of them) are critical and why timing often counts for so much in politics. Most of the earlier work was vague on this point (Pierson n.d.b), although a detailed literature review would be needed to document that claim. The specific characteristics of positive feedback provide a key to making sense of the complex mix of stability and bursts of change that characterize so many political processes. As just discussed, an investigation of increasing returns processes can generate sharper hypotheses, based on more explicit social mechanisms, about the sources of divergent paths and social inertia.

There are, of course, important difficulties with increasing returns arguments. Two require at least brief attention. The first is methodological, and it concerns the difficulty of testing hypotheses based on complex, path dependent arguments (Geddes 1997). The “many variables, few cases” problem is worsened in path dependent arguments, which require one to evaluate sequences of several variables over time.

This need not pose particularly acute problems for studying outcomes when it is possible to generate many cases (e.g., the formation of interest groups). Collective action and the development of actors’ mental maps of politics seem to be promising areas of study. The “few cases, many variables” problem does pose difficulties, however, for increasing returns arguments that operate at a more aggregated level. The need to generate more cases helps explain why comparative politics has always been a field that emphasizes critical junctures (Collier and Collier 1991). Counterfactual analysis is also emerging as an important tool for such studies (Tetlock and Belkin 1996). Furthermore, analysts can use our growing theoretical understanding of path dependent processes to generate more observable implications, for instance, by focusing on intermediate stages in the processes. As Geddes (1997) argues, there are ways to deal with the “small n” problem, but they demand careful research designs. Even careful designs may well be inadequate for anything but fairly simple increasing returns arguments.

A second problem concerns the danger that the increasing returns concept suggests an overly static view of the social world. To take the starkest illustration, Arthur’s Polya urn processes all settle on a particular equilibrium and then essentially stop. Increasing returns processes seem to generate only brief moments of “punctuation” in a largely frozen social landscape. To many, the significance of path dependence is belied by the evident dynamism of social life.

This is a sensible and useful challenge (Thelen 1999). But path dependent analyses need not imply that a particular alternative is permanently locked in following the move onto a self-reinforcing path. Identifying self-reinforcing processes helps us understand why organizational and institutional practices are often extremely persistent—and this is crucial, because these continuities are a striking feature of the social world. Asserting that the social landscape can be permanently frozen hardly is credible, and that is not the claim. Change continues, but it is bounded change—until something erodes or swamps the mechanisms of reproduction that generate continuity. North (1990a, 98–9) summarizes the key point well: “At every step along the way there [are choices]—political and economic—that provide . . . real alternatives. Path dependence is a way to narrow conceptually the choice set and link decision making through time. It is not a story of inevitability in which the past neatly predicts the future.” The claims in path dependent arguments are that previously viable options may be foreclosed in the aftermath of a sustained period of positive feedback, and cumulative commitments on the existing path will often make change difficult and will condition the form in which new branchings will occur.

Indeed, as recently emphasized by Mahoney (n.d.) and Thelen (1999), identifying the particular feedback loops (or “mechanisms of reproduction”) at work often provides key insights into the kinds of events or

---

22 It should be noted, however, that mainstream statistical research will also face difficult challenges if path dependent processes are common, since many quantitative techniques rest on the assumption that they are not (Jackson 1996).
processes that might generate major subsequent change points. Such junctures are usually attributed, often ex post, to "exogenous shocks." We should expect, however, that these change points often occur when new conditions disrupt or overwhelm the specific mechanisms that previously reproduced the existing path.

Increasing returns arguments open up an exciting research agenda for political science. In addition, an understanding of these processes can make another contribution to political scientists: a healthy dose of humility. Since the rise of behavioralism, many political scientists have had lofty aspirations about developing a science of politics, rooted in parsimony and generalization and capable of great predictive power. Despite modest achievements over four decades, these aspirations remain. Setbacks are shrugged off with calls for more time or more sustained application of the proper methods, but the inability to generate powerful generalizations that facilitate prediction remains a puzzle. If the prevalence of increasing returns processes is indeed a defining feature of politics, then we have been looking in the wrong place for an explanation. The main problem lies in the character of the political world itself.

REFERENCES


Marwell, Gerald, and Pamela Oliver. 1993. The Critical Mass in


