Urban Systems and Historical Path-Dependence

W. Brian Arthur

paper number 0012

Morrison Professor of Population Studies and Economics,
Food Research Institute, Encina Hall 311, Stanford, CA 94305

URBAN SYSTEMS AND HISTORICAL PATH-DEPENDENCE

W. Brian Arthur
Stanford University

1. INTRODUCTION

If small events in history had been different, would the pattern of cities we have inherited be different in any significant way? Could different "chance events" in history have created a different formation of urban centers than the one that exists today?

To a great degree, cities form around and depend upon clusters of industry, so that without doing too much injustice to the question we can ask whether the patterns of location of industry follow paths that depend upon history. The German Industry Location School debated this question in the earlier part of this century, but it was never settled conclusively. Von Thünen (1826), the early Weber (1909), Predöhl (1925), Christaller (1933), and Lösch (1944), all tended to see the spatial ordering of industry as preordained—by geographical endowments, shipment possibilities, firms' needs, and the spatial distribution of rents and prices that these induced. In their view, history did not matter: the observed spatial pattern of industry was a unique "solution" to a well-defined spatial economic problem. Therefore, early events in the configuration of an industry could not affect the result. But others, the later Weber, Engländer (1926), Ritschl (1927), and Palander (1935), tended to see industry location as process-dependent, almost geologically stratified, with new industry laid down layer by layer upon inherited, previous locational formations. Again geographical differences and transport possibilities were important, but here the main driving forces were agglomeration economies—the benefits of being close to other firms or to concentrations of industry. In the simplest formulation of this viewpoint (Maruyama 1963), an industry starts off on a uniform, featureless plain; early firms put down by "historical accident" in one or two locations; others are attracted by their presence, and others in turn by *their* presence. The industry ends up clustered in the early-chosen places. But this spatial ordering is not unique: a different set of early events could have steered the locational pattern into quite a different outcome, so that settlement history would be crucial.

These two viewpoints—determinism versus history-dependence, or "necessity" versus "chance"—are echoed in current discussions of how modern industrial clusters have come about. The determinism school, for example, would tend to see the electronics industry in the U.S. as spread over the country, with a substantial part of it in Santa Clara

County in California (Silicon Valley) because that location is close to Pacific sources of supplies, and because it has better access to airports, to skilled labor, and to advances in academic engineering research than elsewhere. Any "small events" are overridden by the "necessity" inherent in the equilibration of spatial economic forces, and Silicon Valley is part of an inevitable result. Historical dependence on the other hand would see Silicon Valley and similar concentrations as largely the outcome of "chance." Certain key persons—the Packards, the Varians, the Shockleys of the industry—happened to set up near Stanford University in the 1940s and 1950s, and the local labor expertise and inter-firm markets they helped to create in Santa Clara County made subsequent location there extremely advantageous for the thousand or so firms that followed them. If these early entrepreneurs had had other predilections, Silicon Valley might well have been somewhere else. In this argument "historical chance" is magnified and preserved in the locational structure that results.

While the historical dependence-agglomeration argument is appealing, it has remained problematical. If history can indeed steer the spatial system down different paths, there are multiple "solutions" to the industry location problem. *Which* of these comes about is indeterminate. In the 1920s, analysts could not cope with this difficulty, and the argument did not gain enough rigor to become completely respectable.

In this paper I will investigate the importance of "chance" (as represented by small events in history) and "necessity" (as represented by determinate economic forces) in determining the pattern of industry location. I will contrast three highly stylized locational models in which small events and economic forces are both present and allowed to interact. In each model an industry is allowed to form, firm by firm, and build up into a locational pattern. In each model I will examine whether historical chance can indeed alter the location pattern that emerges. And I will use the insights gained from the three models to derive some general conditions under which long-run locational patterns may be affected by small historical events.

2. THE EVOLUTION OF LOCATIONAL PATTERNS: THREE MODELS

Model 1. Pure Necessity: Locational Under Independent Preferences

Let us begin with a very simple model indeed of the emergence of an industry location pattern. Starting from zero firms, we allow an industry to form firm by firm, with each new firm that enters deciding "at birth" which of N possible regions (or sites) it will locate in. Once located, each firm will stay put. Firms in this industry are not all alike; there are of I different types. The net present value or payoff to a firm of type i for locating in region j is $\Pi^i{}_j$, and each firm choosing selects the location with the highest return for its type. In this model firms are independent: the presence or absence of other firms does not affect what they can earn in each region.

We now inject a small element of "chance" by assuming that the particular historical circumstances that lead to the next firm's being of a particular type are unknown. We do know, however, that a firm of type I will occur next with probability p_i. The question is: what pattern of industrial settlement will emerge in this model, and can it be affected by a different sequence of historical events in the formation of the industry?

It is not difficult to work out the probability that at any time of choice region j will be chosen. This is simply the probability, q_j, that the newest firm is of a type which has its highest payoff in region j which is given by $q_j = \sum p_k$ for $k \in K$, where K is the set of firm types that prefer j. Repeating this calculation for each of the N regions, we have a set of probabilities of choice $q = (q_1, q_2, \ldots, q_N)$, that are constant no matter what the current pattern of location is. Starting from zero firms in any region, concentrations of the industry in the various regions will fluctuate, considerably at first. But the strong law of large numbers tells us that as the industry grows, the proportions of it in the N regions must settle down to the expectation of an addition being made to each region. That is, regions' shares of the industry must converge to the constant vector q. In this simple model then, even though well-defined "chance historical events" are present, a unique, predetermined locational pattern emerges and persists.

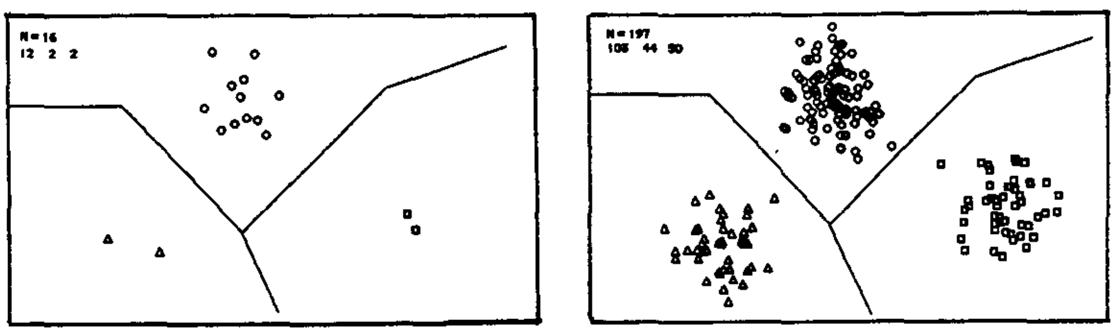

Figure 1. A Three-Region Example of the Independent Preferences Location Model.

Figure 1 shows a simple three-region simulation of this process, with three possible firm types that prefer (clockwise from the top) region 1,region 2, and region 3 respectively, with probabilities of occurrence 0.5, 0.25, and 0.25. After 16 firms have located, the regions' shares of the industry are 0.75, 0.125, 0.125 respectively—not yet close to the long-run predicted pattern. After 199 firms have located however, the shares are 0.528, 0.221, 0.251—much closer to the predetermined theoretical long-run shares.

In this model, chance events, represented as randomness in the sequence of firm types that enter the industry, are important early on. But they are progressively averaged away to become dominated by the economic forces represented by firms' payoffs in each region. Different sequences of firm types caused by different historical events would, with probability one, steer the system into the same locational pattern. Here historical chance cannot affect the outcome. Necessity dominates.

Model 2. Pure Chance: Location by Spinoff.

We now assume a quite different mechanism driving the regional formation of an industry—one in which chance events become all-important. Once again the industry builds up firm by firm, starting with some set of initial firms, one per region, say. This time new firms are added by "spinning off" from parent firms one at a time. (David Cohen (1984) has shown that such spin-offs have been the dominant "birth mechanism" in the

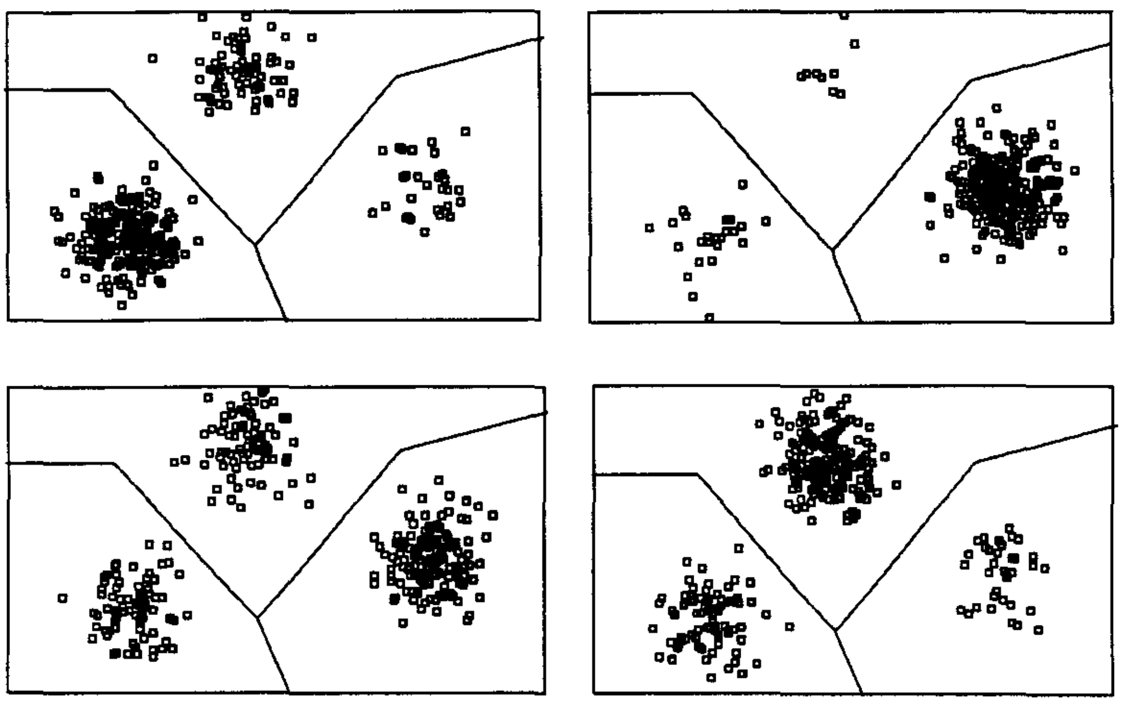

Figure 2. Four Realizations of Location-by-Spinoff.

U.S. electronics industry.) We assume that each new firm stays in its parent location, and that any existing firm is as likely to spin off a new firm as any other. With this mechanism we have a different source of "chance historical events:" the sequence in which firms spin off daughter firms.

It is easy to see that in this case firms are added incrementally to regions with probabilities exactly equal to the proportions of firms in each region at that time. This random process, in which unit increments are added one at a time to one of N categories with probabilities equal to current proportions in each category, is known in probability theory as a *Polya process*. We can use this fact to examine the long-term locational patterns that might emerge. From Polya theory we know that once again the industry will settle into a locational pattern (with probability one) that has unchanging proportions of the industry in each region. But although this vector of proportions settles down and becomes constant, surprisingly it settles to a constant vector that is *selected randomly* from a uniform distribution over all possible shares that sum to one. This means that each time this spin-off locational process is "rerun" under different historical events (in this case a different sequence of firms spinning-off) it will in all likelihood settle into a different pattern. We could generate a representative outcome by placing N-1 points on the unit interval at random, and cutting at these points to obtain N "shares" of the unit interval.

Figure 2 shows four realizations of this location-by-spin-off mechanism starting from the same three original firms in a three-region case. Each of the four "reruns" has settled into a pattern that will change but little in regional shares with the addition of further firms. But each pattern is different from the others. In this model industry location is highly path-dependent. Although we *can* predict that the locational pattern of industry will indeed settle down to constant proportions, we cannot predict what proportions it will settle into. Any given outcome—any vector of proportions that sum to one—is as likely as any other. "History" in the shape of the early random sequence of spin-offs becomes the sole determining factor of the regional pattern of industry. In this model "chance" dominates completely.

Model 3. Chance and Necessity: Location under Agglomeration Economies.

Firms that are not tied to raw-material localities and that do not compete for local customers are often attracted by the presence of other firms in a region. More densely settled regions offer better infrastructure, deeper labor markets (David 1984), more specialized legal and financial services, better local availability of inventory and parts, and more opportunity to do business face-to-face. For our third model we go back to model 1 and extend it by supposing that new firms gain additional benefits from local agglomerations of firms.

Suppose now that the net present value or payoff to a firm of type i for locating in region j is $\Pi^i_j + g(y_j)$, where the "geographical benefits," Π^i_j, are enhanced by additional "agglomeration benefits," $g(y_j)$, from the presence of y_j firms already located in that region. We can recalculate the probability that region j is chosen next, given that $y_1, ..., y_N$, firms are currently in regions 1 through N, once again as $q_j = \sum p_k$ for $k \in K$, where K is now the set of firm types for which $\Pi^i_j + g(y_j) > \Pi^i_m + g(y_m)$ for all regions $m \neq j$. Notice that in this case the probability that region j is chosen is a function of the number of firms in each region at the time of choice.

Starting from zero firms in the regions, once again we can allow the industry to grow firm by firm, with the appearance of firm-types subject to known probabilities as in model 1. Again the pattern of location of the industry will fluctuate somewhat, but in this model if a region by a combination of luck and geographical attractiveness gets ahead in numbers of firms, its position is enhanced. We can show (see Arthur 1986 for proof) that if agglomeration benefits increase without ceiling as firms are added to a region (that is, if the function g is monotonically increasing without upper bound) then eventually (with probability one) one of the regions will gain enough firms to offer sufficient locational advantages to shut the other regions out in all subsequent locational choices.[1] From then on, each entering firm in the industry will choose this region, and this region's share of the industry will tend to 100 percent, with the others' shares tending to 0 percent.

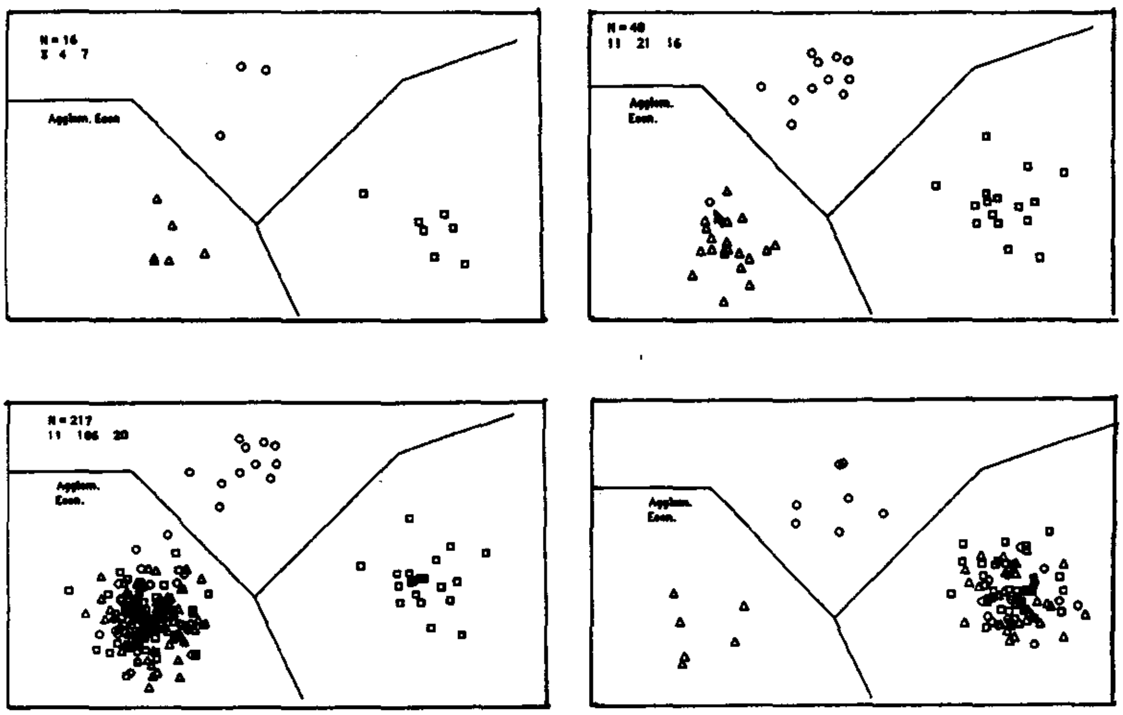

Figure 3. Two Realizations of a Locational Process with Agglomeration Economies.

Figure 3 shows two realizations of a three-region example with agglomeration economies. The first three panels show the build-up of firms, with geographical preferences dominating in panel 1, but with region three in panel 2 by good fortune in the sequence of arrival of firm-types just gaining enough firms to cause another firm-type to favor it instead of its pure geographical preference. In panel 3 region three has come to dominate the entire industry in a Silicon Valley-like cluster. Panel 4 show the outcome of an alternatve run. Here the industry is locked-in to region two.

In this unbounded agglomeration economies model, monopoly of the industry by a single region must occur. But *which* region achieves this "Silicon Valley" locational monopoly is subject to historical luck in the sequence of firm types choosing. Chance, of course, is not the only factor here. Regions that are geographically attractive to many firm types—regions that offer high economic benefits—will have a higher probability of being selected early on. And this will make them more likely to become the single region that dominates the industry. To use an analogy borrowed from genetics, chance events act to "select" the pattern that becomes "fixed"; but regions that are economically attractive enjoy "selectional advantage," with correspondingly higher probabilities of gaining dominance. In this third model the long-run locational pattern is due both to chance and necessity.

3. PATH-DEPENDENCE AND CONVEXITY

Each of our three stylized industry location models includes both determinate economic forces and some source of chance events. But each behaves differently. Determinate forces, or historical chance, or a mixture of the two, are in turn responsible for the long-run pattern of industry settlement that emerges.

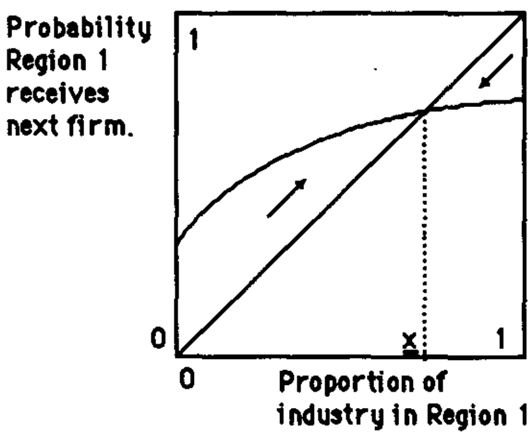

Figure 4. Proportion-to Probability Mapping.[a]
[a]Arrows indicate expected motions

To explain these results, and to provide some precise conditions under which historical chance can be important, it is useful to introduce a general framework that encompasses all three models (as well as many others). In this general framework, suppose there are N regions, and that industry locates, one firm at a time, starting from a given number of firms in each region. Different economic forces, different sources of chance events, and different mechanisms of locational choice would be possible within this framework, but we do not need to know these. What we do need to know are the probabilities that region 1, region 2, ..., region N will be chosen next, as a *function* of current regional shares of the industry $x_1, x_2, ..., x_N$. Plotting this function (as in Figure

4 for the 2-region case) we might expect that where the probability of a region's receiving the next firm exceeds its current proportion of the industry, it would tend to increase in proportion; and where the probability is less than current proportion, it would tend to decrease in proportion. Moreover, as firms are added, each new addition changes proportions or shares by an ever smaller magnitude. Therefore proportions should settle down, and fluctuations in proportions should die away. In the long run then, we might expect that regions' proportions (the industry's location pattern) ought to converge to a point—to a vector of locational shares—where proportions equal probabilities, and to one that expected motions lead toward (point $\underline{x}$ in Figure 4). That is, it ought to end up at a

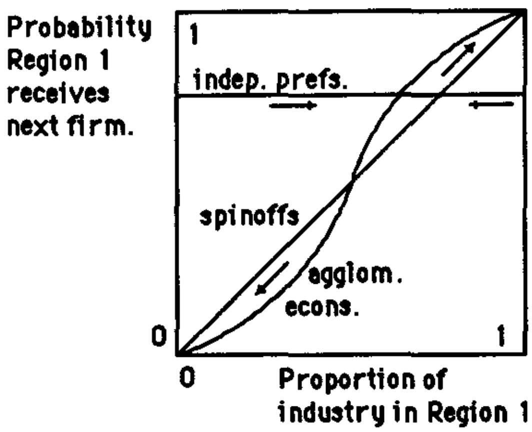

Figure 5. Probability Mappings for the Three Models.

stable fixed point of this proportions-to-probabilities function. It takes powerful theoretical machinery to prove this conjecture, but it turns out to hold under unrestrictive technical conditions (see Hill, Lane, and Sudderth 1980, and Arthur, Ermoliev, and Kaniovski 1983, 1986a, 1986b).[2] Further, and significantly for us, where there are multiple stable fixed points, each of these would be candidates for the long-run locational pattern, with different sequences of chance events steering the process toward one of the multiple candidates.

We can now see what happened in our three locational models (Figure 5). The first model, "independent-preferences," has constant probabilities of choice, and thus a single fixed point. Therefore it has a unique, predetermined outcome. The second model, "spin-off," with probabilities equal to proportions, has *every* point a fixed point; so that "chance" could drive this locational process to any outcome. The third model, "agglomeration economies," has 0 and 1 as two candidate stable fixed points. Thus the outcome is not fully predetermined and one of the candidate solutions is "chosen" by accumulation of chance events.

When does history count in the determination of industry location patterns? We can now answer this question, at least for the broad class of models that fit our general framework. "History"—the small elements outside our economic model, that we must treat as random—becomes the determining factor when there are multiple solutions or multiple fixed points in the proportions-to-probabilities mapping. More intuitively, history counts when expected motions of regions' shares do not always lead the locational process toward the same share.

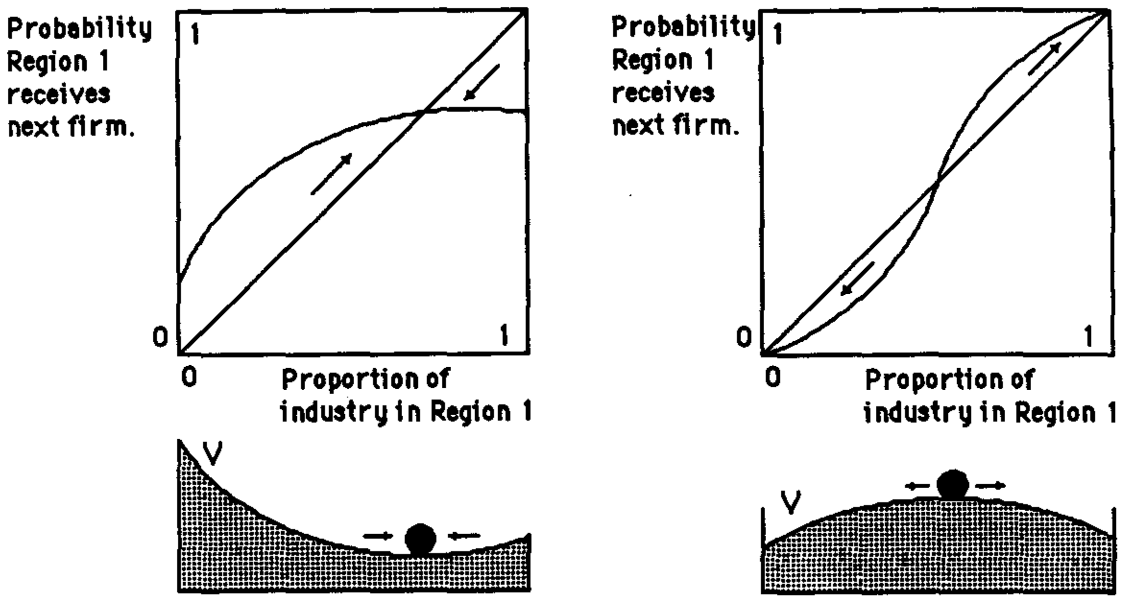

Figure 6. Convex and Nonconvex Potential Functions.

It is useful to associate with each probability function a *potential function* V whose downhill gradient equals the expected motion of regions' shares.[3] (See Figure 6.) Intuitively, we can think of the process as behaving like a particle attracted by gravity to the lowest points on the potential, subject to random fluctuations that die away. If this potential function is convex (looking upward at it) it has a unique minimum, and therefore the locational process that corresponds to it has a unique determinate outcome which expected motions lead toward and which historical chance cannot influence. If, on the other hand, it is non-convex, it must have two or more minima, with a corresponding split in expected motions, and with "historical chance" determining which of these is ultimately selected.

To establish non-convexity, all we need is the existence of at least one unstable point, a "watershed" share of the industry, above which the region with this share exerts enough attraction to increase its share, below which it tends to lose its share. But in a way, this is another definition of the presence of agglomeration economies: If above a certain density of settlement, a region tends to attract further density, and if below it tends to lose density, there must be some agglomeration mechanism present. The underlying system will then be non-convex, and history will count.

4. CONCLUDING REMARKS

Whether small events in history matter in determining the pattern of spatial or regional settlement in the economy reduces, strangely enough, to a question of topology. It reduces to whether the underlying structure of locational forces guiding the locational pattern as it forms is convex or non-convex. And for this structure to be non-convex, so that history will matter, there must be some mechanism of agglomeration present.

Our models were highly stylized. They considered populations of firms, not people; they assumed that firms lived forever and never moved; and they dealt with the formation of one industry only over time, not several. Nevertheless, even if the mechanisms creating urban systems in the past and present are a great deal more complex, it is still likely that a *mixture* of economic determinism and historical chance—not either alone—has formed the spatial patterns we observe. Certain firms, like steel manufacturers, do need to be near sources of raw materials, and for them spatial economic necessity dominates historical chance. Certain firms, like gasoline distributors, do need to be separated from their competitors in the same industry, and for them the necessity to spread apart again dominates historical chance. But most firms need to be near other firms, if not in their own industry then in other industries that act as their suppliers of parts, machinery, and services, or as consumers of their products and services. And for this reason they are attracted to existing and growing agglomerations. After all, it is this need of firms to be near other firms that causes cities—agglomerative clusters—to exist at all.

Thus it is highly likely that the system of cities we have inherited is only partly the result of industries' geographical needs, raw material locations, the presence of natural harbors, and transportation costs. It is also the result of where immigrants with certain skills landed, where early settlers met to market goods, where wagon trains stopped for the night, where banking services happened to set up, and where politics dictated that canals and roads and railroads be built.

We cannot explain the observed pattern of cities by economic determinism alone, without reference to chance events, coincidences, and circumstances in the past. And without knowledge of chance events, coincidences, and circumstances yet to come, we cannot predict with accuracy the shape of urban systems in the future.

NOTES

1. In the case where g is bounded, several locations can share the industry in the long run. But again typically there are multiple possible outcomes, so that chance events matter here too (see Arthur 1986,1985).
2. The set of fixed points needs to have a finite number of components. Where the proportions-to-probabilities function itself changes with the number of firms located, as

in the agglomeration case, the theorem applies to the limiting function fo these changing functions, providing it exists. (See Arthur, Ermoliev and Kaniovski (1986a,1986b).

3. For dimension N > 2 a potential function may not exist. This would be the case if there were cycles or more exotic attractors than the single-point ones that are considered here.

REFERENCES

ARTHUR, W.B. (1985), "Competing Technologies and Lock-In by Historical Small Events: The Dynamics of Choice under Increasing Returns", C.E.P.R. Paper 43. Stanford.

ARTHUR, W. B. (1986), "Industry Location Patterns and the Importance of History," C.E.P.R. Paper 84. Stanford.

ARTHUR, W.B., ERMOLIEV, YU. M. and KANIOVSKI, YU. M. (1983), "A Generalized Urn Problem and Its Applications," *Cybernetics* 19, 61-71.

ARTHUR, W.B., ERMOLIEV, YU. M., and KANIOVSKI, YU. M. (1986a), "Strong Laws for a Class of Path-Dependent Urn Processes", in *Procs. International Conf. on Stochastic Optimization, Kiev 1984,* Arkin, Shiryayev, and Wets (eds.), Springer: Lecture Notes in Control and Info. Sciences 81.

ARTHUR, W.B., ERMOLIEV, YU. M., and KANIOVSKI, YU. M. (1986b), "Path-Dependent Processes and the Emergence of Macro-Structure", forthcoming, *European Journal of Operations Research.*

CHRISTALLER, W. (1933), *Central Places in Southern Germany.* (1966) Prentice-Hall.

COHEN, D.L. (1984), "Locational Patterns in the Electronics Industry: A Survey." Mimeo. Stanford.

DAVID, P. (1984), "High Technology Centers and the Economics of Locational Tournaments", Mimeo. Stanford.

ENGLÄNDER, O. (1926), "Kritisches and Positives zu einer allgemeinen reinen Lehre vom Standort", *Zeitschrift für Volkswirtschaft und Sozialpolitik,* Neue Folge, 5.

HILL, B., LANE, D., and SUDDERTH W. (1980), " Strong Convergence for a Class of Urn Schemes", *Annals of Probability,* 8, 214-226.

LÖSCH, A. (1941), *The Economics of Location.* Yale University Press (1954).

MARUYAMA, M. (1963) "The Second Cybernetics: Deviation Amplifying Mutual Causal Processes" *American Scientist* 51, 164-179.

PALANDER,T. (1935), *Beiträge zur Standortstheorie,* Almqvist and Wicksell.

PREDÖHL, A. (1925), "Das Standortsproblem in der Wirtschaftslehre", *Weltwirtschaftliches Archiv,* 21, 294-331.

RITSCHL, H. (1927), "Reine und historische Dynamik des Standortes der Erzeugungszweige", *Schmollers Jahrbuch,* 51, 813-70.

THÜNEN, J. H. VON, (1826), *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie,* Hamburg.

WEBER, A. (1909), *Theory of the Location of Industries,* (1929) Univ. of Chicago Press.