Path Dependence in Spatial Networks: The Standardization of Railway Track Gauge

Douglas J. Puffert

Institute for Economic History, University of Munich
Ludwigstr. 33 / IV, 80539 Munich, Germany
Telephone +49-89-2180-3035 Fax +49-89-339233
douglas.puffert@econhist.vwl.uni-muenchen.de

This version: July 2001

Abstract

This article investigates the historical emergence of regional standard railway track gauges in light of a model of the interaction of agents' choices within a spatial network. Contingent events, reinforced by positive feedbacks, determined both particular standards and the geographic extent of standardization in Britain, Continental Europe, North America, and Australia. The model, solved using numerical simulation, shows the allocation process to be path dependent. Monte Carlo experiments demonstrate how the distribution of possible outcomes varies with historically varying systematic factors. Both history and an extension to the model demonstrate the role of externality-internalizing behavior in resolving diversity.

Keywords: railways, networks, path dependence, technical standards, cellular automata
JEL classifications: O33, N70, C63

Path Dependence in Spatial Networks:
The Standardization of Railway Track Gauge

One of the most vigorously disputed issues among economic historians is the importance of path-dependent processes of allocation in determining features of the economy. Over 250 comments on path dependence were posted to an economic historians' email discussion group over a recent five-year period, about three times the number addressing the next most common topic.[1]

A path-dependent economic process is one in which specific contingent events—and not just fundamental determinative factors like technology, preferences, factor endowments, and institutions—have a persistent effect on the subsequent course of allocation. Paul A. David (1985, 1993) and W. Brian Arthur (1989, 1994) have contrasted the multiple potential outcomes of path-dependent processes to the unique, necessarily efficient outcomes typically predicted in standard neoclassical models. Critics of the concept of path dependence, led by S.J. Liebowitz and Stephen E. Margolis (1990, 1994, 1995), have emphasized the view that forward-looking, profit-seeking agents steer allocation processes to the best outcomes possible given the constraints of foresight and transactions costs. At stake in this dispute, as both sides note, is the sense in which "history matters" for explaining the present economy.

The present article offers a partial reconciliation of these differing views. It shows both empirically and in a model how path dependence arises precisely when either foresight (or information) is lacking or else externalities prevent the sorts of behavior that could direct an allocation process toward a unique and optimal outcome. Although Liebowitz and Margolis (1995) dismiss path dependence under such conditions as not mattering and as offering no challenge to "the neoclassical model of relentlessly rational behavior leading to efficient, and therefore predictable, outcomes," I show that such path dependence indeed affects economic structure and efficiency, and that it explains features of the economy that are not explained by what Liebowitz and Margolis call the "neoclassical model."

I examine these issues using the historical selection of regional standards for railway track gauge—the distance between a pair of rails.[2] Railway companies or administrations that share a common gauge can much more easily exchange traffic, resulting in lower costs, improved service, and greater profits. As a result, positive “network” externalities (Katz and Shapiro, 1985, 1994) produce positive feedbacks among choices of gauge by different agents.

Nevertheless, in many parts of the world diversity in gauge arose and, often, persists to this day (table 1). Australia and Argentina each have three different regional-standard gauges, although this is recognized as a costly hindrance to national commerce. India, Japan, Chile, and several other countries each make extensive use of two gauges. “Breaks of gauge” hinder through-service across numerous international borders, including that of France with Spain and most external borders of the former Russian and Soviet empires.

[Table 1 here]

To be sure, much costly diversity has been eliminated. The United States and Canada had six gauges in widespread use until the 1880s. Now only a few relic tourist lines use variant gauges. Britain’s extensive Great Western Railway system used a variant gauge for over 50 years before completing its conversion to the gauge of neighboring systems in 1892. Similarly, the original gauges of the Netherlands, the earlier German state of Baden, and much of Norway gave way to the common standard that emerged in most of western and central Europe. In recent decades, Australia and India have made substantial progress in reducing their diversity of gauge.

What explains the emergence, persistence, and in some cases resolution of diversity in regional railway-network standards? To investigate this question, I first examine certain details of the worldwide history of track gauge. I investigate the incentives of railway builders and operators, seeking to determine how individual and collective choices of gauge depended both on contingent past events and on systematic tendencies to optimize the system-level outcome. I find evidence both for positive feedbacks, particularly in the earlier stages of the aggregate gauge-selection process, and for systematic rationalization of outcomes, particularly in later stages of the process. In no case did the process entirely break free of early contingent choices and events. “Founder” effects have persisted, most notably in the worldwide predominance to this day of the gauge that engineer George Stephenson transferred from a primitive mining tramway to the Liverpool and Manchester Railway. Nevertheless, there appear to be systematic reasons why regional standards have given way to larger-scale standardization in some countries and continents but not in others.

To examine these reasons further, I investigate the underlying dynamic structure of the gauge selection process by developing a modeling framework that considers the interaction of agents' choices within a spatial network. I show how both the extent of overall diversity and the particular techniques that emerge as regional standards may depend on specific contingent events. I also show how the distribution of possible outcomes may vary with variations in fundamental factors and network structure—features that varied among different historical contexts. In an extension to the modeling framework, I show how systematic optimization through the internalization of externalities, as argued by Liebowitz and Margolis (1994, 1995), can in many but not all cases rationalize and improve the outcome of a path-dependent process. I consider how each of these results affect the interpretation of the history.

The purpose of this article is to investigate how regional standard gauges have arisen, persisted, and in some cases been superseded. The chief economic issue at stake has been the extent of standardization and diversity, not the selection of suboptimal gauges. At least some gauges in use are suboptimal, as most railway engineers hold that the optimal gauge for most applications is somewhat broader than the common Stephenson gauge of 4 feet 8.5 inches (4'8.5"—1435 mm.), although they do not consider the extent of inefficiency to be great.[3] However, I do not argue that early choices of specific gauges were "wrong" given the technical conditions and understanding of their time, or that markets, other institutions, or entrepreneurs have "failed" in continuing to use them.

The case of railway track gauge offers the opportunity to consider the empirical relevance both of path dependence in general as an explanation of some economic allocation processes and of a particular class of models sometimes used to explain path dependence. The case has an advantage over such disputed cases as the QWERTY typewriter keyboard (David, 1985; Liebowitz and Margolis, 1990) in that there have been numerous well documented local realizations of the process. This makes it far easier to differentiate between the effects of contingent events with positive feedbacks (David's emphasis) and systematic, forward-looking optimizing behavior (Liebowitz' and Margolis' emphasis)—and to consider how these factors interact. Furthermore, this study considers a concretely spatial path-dependent process, an application that David (1993) has cited as a needed addition to the literature.

I. Episodes in the History of Gauge

In reviewing the history of gauge, it quickly becomes evident that contingent events and positive feedbacks played a major role in deciding which particular gauges became the local standards of particular regions, although agents' choices were certainly efforts to optimize. Contingent events also often decided the number of different gauges introduced to different parts of regions within which traffic exchanges would later make a common gauge desirable. As the costs of diversity increased, systematic incentives and optimizing behavior greatly reduced this diversity, but in some cases early contingent diversity persists to the present.[4]

Britain

Great Britain was the first country to develop modern railways, and events there had a world-wide impact. A large variety of gauges were used for the primitive railways that developed in mining districts during the late eighteenth century, including 4'8" (1422 mm.) on a small group of lines that brought coal to the river Tyne near Newcastle. It was there, however, that the gifted mechanical engineer George Stephenson performed early experiments with steam locomotion during the 1810s. In recognition of his broad abilities, Stephenson was asked to build the two railways that together introduced a new era of construction and operating practice, the Stockton and Darlington Railway, opened in 1825, and the Liverpool and Manchester (L&M) Railway, opened in 1830. The L&M was the first railway designed exclusively for steam locomotion and the first to rely exclusively on commercial and passenger rather than mining traffic. Nevertheless, Stephenson used the same 4'8" gauge as before—except for adding half an inch (13 mm.) between the rails to allow for more space between rails and wheel flanges.

Stephenson gave no particular thought to the question of optimal gauge but rather simply followed precedent. As Stephenson's son Robert later told a parliamentary commission, his father did not "propose" the gauge but rather "adopted" what was already in use in his home region (Great Britain, 1846, Minutes ¶7). Stephenson's friend and biographer Samuel Smiles (1868, p. 424) wrote that the gauge "was not fixed after any scientific theory, but adopted simply because its use had already been established." By contrast, Stephenson's rivals for the contract to build the L&M proposed an unprecedentedly broad gauge, 5'6" (1676 mm.), as a reflection of what they regarded as a new engineering problem (Carlson, 1969). If that rival team or someone else had built the L&M, or if Stephenson had gotten his early experience elsewhere, then the L&M's gauge would almost surely have been different.

Stephenson's involvement with the L&M was the most crucial contingent event in the history of track gauge. His choice of gauge generated positive feedbacks through several mechanisms. First, the Stephenson gauge was adopted for the sake of traffic exchange by an expanding network of lines that soon branched out from the L&M eastward into Yorkshire and southward to Birmingham and London. The gauge diffused directly to still other regions in Britain because Stephenson himself and his protégés used it, because other engineers accepted it as representing best practice, and because specification of the gauge was briefly a standard feature of parliamentary acts to authorize new railways.

Beginning in the mid-1830s, however, some British locomotive builders found their ability to develop increasingly powerful, easily maintained engines constrained by the 4'8.5" gauge, while certain civil engineers expected that a broader gauge would promote improved stability, smoothness of ride, speed, and capacity. As a result, a few short lines adopted 5'0" (1524 mm.) and 5'6" for what they initially expected to be isolated local networks. When the lines were reached by the expanding Stephenson-gauge network, they converted immediately.

A much more important source of contingent diversity was Britain's second great railway engineer, Isambard Kingdom Brunel, builder of the extensive Great Western Railway (GWR) system west of London. More than any of his peers, Brunel was convinced that a quite broad gauge—fully 7'0" (2134 mm.)—was needed for the full development of railway technology. He argued, furthermore, that the GWR system would form a self-contained railway district, with little need to exchange traffic externally and thus unhindered by breaks of gauge. Many have interpreted this as an effort to use gauge to monopolize the region's traffic.

Brunel was soon proved wrong on the importance of breaks of gauge. Not only did they become a major public issue in 1845 as an "evil," leading to a parliamentary investigation and then legal restrictions to the spread of the gauge beyond the GWR, but they proved costly to the GWR system itself, both in transshipment costs and loss of traffic. Nevertheless, as the GWR system grew to encompass an increasing number of Stephenson-gauge routes beyond its original boundaries, it was able to manage the diversity in a relatively rational, efficient way, in part by using mixed gauges—three-rail track—on trunk routes serving both gauges. From 1868 to 1892, the GWR progressively converted to the Stephenson gauge.

Continental Europe

Belgium, France, Austria, and several of the then independent German and Italian states adopted the Stephenson gauge during the mid- to late 1830s. Stephenson himself introduced the gauge to Belgium, his protogés or other British engineers did so in several countries, and in other places local engineers either accepted the gauge as one element of current best practice or else simply fitted their track to British locomotives. This common influence greatly limited the amount of diversity that might have developed. Some of the German states apparently followed the prior choices of others, as an integrated German railway network was part of the pan-German economic program promoted by Friedrich List. Prussia was interested in a common-gauge link to France, but otherwise there is little evidence that choices of gauge were influenced initially by the desire to develop an integrated continental network. Governments did ensure, however, that domestic railways adopted a common gauge.

The lack of interest in international standardization is clearly evident in the adoption of broad gauges during the late 1830s and 1840s by the Netherlands (1945 mm.), the German grand duchy of Baden (1600 mm.), Russia (1524 mm.), and Spain (1672 mm.). Following much of contemporary British opinion and practice, the government-commissioned engineers (local except in Russia) who selected each of these gauges saw them as embodying a more advanced railway technology. Contrary to a common belief, Russia's gauge was not chosen as a defensive military measure (Haywood, 1969). With the probable exception of Baden, these countries did not foresee that railways would soon begin to displace water transport in international commerce; Baden sought to have neighboring countries adopt the same gauge.

Sooner or later, all of these countries came to regret their choices. The Netherlands found itself losing entrepôt trade to Belgium due to the latter country's well-developed railway system and common-gauge connections to Germany, and the Belgian network's expansion over the border threatened to draw domestic trade away from the Netherlands' own commercial centers. When Prussia expressed interest in a common-gauge connection to Amsterdam and Rotterdam in the early 1850s, the Netherlands converted. In the case of Baden, neighboring states preferred to adopt the much more common Stephenson gauge, and Baden itself converted in 1854-55 as it initiated a new wave of construction.

The variant gauges of Russia and Spain remain to this day, as these more peripheral countries had little exchange of traffic with the core of Europe until their common-gauge networks—and potential conversion costs—had grown relatively large. Russia's choice began to be costly during the 1860s, when the main Russian network advanced into Russian-ruled Poland, which had adopted the Stephenson gauge in 1839 in order to gain an outlet for international commerce through Austria to Trieste as an alternative to the Prussian-controlled mouth of the Vistula. Spain's (and Portugal's) choice mattered relatively little until the recent integration of Spain and Portugal into the economy of the European Union. An estimated cost of (U.S.) $5 billion has prevented conversion, but Spain is reducing the cost of hoped-for future conversion by introducing dual-gauge prefabricated concrete cross-ties during routine track maintenance. Spain adopted the Stephenson gauge for its high-speed train lines for the sake of a future connection to France's TGV, at the cost of an awkward diversity of gauge within the country today.

In 1862, Norway pioneered the development of narrow-gauge railways. By this time the main difficulties in locomotive design that had previously favored broad gauges had been resolved, and it became possible to take advantage of the ability of narrow gauges to make sharper curves, following the contours of rugged or mountainous landscape and reducing the need for costly tunnels, cuttings, bridges, and embankments. The narrow gauge was confined to lines north and west of Oslo that were expected to be used primarily for local traffic, but a new focus after 1900 on developing a nationally and internationally integrated network led to the gradual conversion and upgrading of these lines.

Beginning in the 1870s, narrow gauges were widely used for lines in the Alps and other mountains as well as for extensive systems of light railways used to bring agricultural produce to market in several parts of Europe. Many of the former lines are still in service at their original gauges, but the latter have been replaced by modern road transport.

North America

Builders of the earliest North American railways also regarded the Stephenson gauge as best practice, but they interpreted this practice loosely, introducing gauges of 4’10” (1473 mm.) and 5’0”, as well as 4’8.5”, between 1830 and 1832. During these earliest years, railways were seen as inferior substitutes for waterways, used for routes where canal construction was impractical. They served strictly local purposes, and their builders did not foresee the later importance of a precise common standard. The gauge of 4’8.5” was introduced by far the most often in new regions, including by the great majority of the scattered early lines in the southeastern United States. Nevertheless, the major network spanning that region happened to develop as a series of lines connecting to the original 5’0”-gauge railway, and this became the regional standard gauge. Similarly, the network of the eastern Midwest (chiefly Ohio) expanded from a single 4’10” line, forming a barrier between Stephenson-gauge regions to the east and west. The introduction of 4’10” to Ohio resulted from the happenstance purchase of a surplus locomotive from a different region “off the shelf.” Otherwise there is no clear case where equipment supply determined gauge in North America, as manufacturers supplied all major gauges and also built to order.

From 1838 to the early 1850s, builders also introduced broad gauges of 6’0” (1828 mm.) and 5’6” for what they thought would be self-contained systems. Indeed, in two cases, these gauges were chosen not only for their presumed technical superiority but also precisely because they differed, for the purpose of controlling regional traffic. However, as interregional traffic grew greatly in importance, the variant gauges served much more to keep traffic out of the systems than to keep traffic in.

As a result of these early events, nine different common-gauge regions emerged by the 1860s, including three separated regions using the Stephenson gauge. This diversity was resolved over the period 1866-1886 as a result of three developments: the strong growth in demand for interregional transport, including for the shipment of Midwestern grain to the seaboard; the growth of cooperation among separately owned lines; and the consolidation of interregional trunkline systems under common ownership. The first development increased the level of potential network integration benefits (i.e., network externalities) relative to conversion costs; the others facilitated the internalization of externalities.

In 1866, the Stephenson-gauge New York Central and Michigan Central Railroads offered side payments to the intervening Great Western Railway of southwestern Ontario to lay a Stephenson-gauge third rail over its 5’6”-gauge route, creating the first “bridge” route linking separated Stephenson-gauge regions. In 1869, the Pennsylvania Railroad (PRR) took long-term leases of three trunk routes across Ohio, narrowing their gauge and linking the mid-Atlantic Stephenson-gauge region to the Midwest.[5] Other railways in Canada and Ohio then changed their gauge individually, as the first conversions made it more profitable for other lines to change their gauge as well. As a large, core common-gauge network emerged, other variant-gauge regions converted in order to gain the benefits of network integration.

The last region to convert was the 5’0”-gauge southeastern United States. Several lines on the periphery of this region converted individually, including the interregional Illinois Central Railroad’s route to New Orleans. Thereafter, the 14 major remaining lines made a coordinated decision to convert together, thus preserving their mutual links while integrating into the emerging continental network.

Even as the early diversity was being resolved, a “narrow-gauge fever,” based largely on unrealistic expectations of cost savings, led to the construction of over 20,000 miles of 3’0” (914 mm.) and 3’6” track. The costs of breaks of gauge, together with the financial failure of a “National Narrow-Gauge Trunk” in 1883, led to a sharp decline in new construction, but some local systems remained in service for several decades (Hilton, 1990).

Australia

Australia offers an example of institutional failure in the emergence and persistence of gauge diversity. In the early 1850s, the colony of New South Wales first chose 5’3” (1600 mm.) as its gauge and persuaded Victoria and South Australia to adopt the same measure. Then New South Wales changed its chief engineer and followed his recommendation to change the planned gauge to 4’8.5”. Victoria, which had already ordered equipment from Britain for the broader gauge, appealed to the British colonial administration to intervene, but the latter applied the principle of *laissez faire* in refusing. The estimated cost of remedying the resulting diversity rose, as equipment was purchased and track was laid, from £15,000-£20,000 in 1853, when breaks of gauge were a distant prospect, to £2.4 million in 1897 and £12.1 million in 1913, when they were becoming costly. Efforts to resolve the diversity were long hindered by disputes over how the separate government-owned systems should divide the costs (Harding, 1958). From 1957 to 1982, the national government sponsored new Stephenson-gauge routes to form a nationwide system linking state capitals. During the 1990s, Victoria and South Australia converted their most major routes, and more conversions are expected to follow.

Rest of the World

The patterns of gauge selection in Latin America, Africa, and Asia can be addressed here only in very broad strokes. Regions where railways were introduced by the 1860s adopted either the Stephenson gauge or broader gauges; regions where railways were introduced later adopted the Stephenson gauge or narrower. Because railway builders differed in their preferred gauges, diversity emerged as local common-gauge networks of different gauges came into contact. Less of this diversity was resolved than in Europe and North America, in large part due to lower demand for interregional and international transport.

Japan is noteworthy for introducing new diversity in recent times. Finding its 3’6” gauge unsuitable for high-speed service, Japan introduced the Stephenson gauge in 1964 for its Shinkansen “bullet”-train system. Since 1990, this diversity has hampered efforts to expand high-speed service and integrate the Shinkansen system into the rest of Japan’s network. Some short sections of track have been converted to the broader gauge or to mixed gauges.

Common Elements to the History

Among the common elements to different regional histories was the mix of incentives governing choice of gauge. First, railway builders, operators, and in some cases regulators have had preferences over specific gauges, based on perceptions of the technical performance characteristics of different gauges. Second, agents have nearly always valued compatability with neighboring railways, adopting established gauges where they existed. The first incentive has been a source of variation in gauge practice; the second incentive a source of commonality of practice through positive feedbacks among the choices of different agents.

Historically, an interest in compatibility was often relatively weak in the early years of railways. Railway builders did not foresee the future value of long-distance railway transport, and thus they placed little value on compatibility with previous lines, except for those nearby. As time went on, railway builders placed an increasing value on compatibility, but in some cases they also placed increasing value on particular variant gauges—generally broad gauges from the late 1830s to the 1860s and narrow gauges from the 1860s to about 1900. In rare cases, variant gauges were chosen partly for the purpose of controlling regional traffic. Equipment supply—particularly of locomotives—seems to have affected only a few choices in Europe and one in North America, as suppliers offered equipment for all the usual gauges and also built to order.[6]

Early choices of gauge were generally made by individual local railway companies or governments, with little regard for the effects of their choices on others. Later, cooperation and the formation of interregional railway systems led to increased coordination of choices, often facilitating the resolution of early diversity.

II. The Modeling Framework

I seek to capture these incentives in a modeling framework that sheds further light on the underlying dynamic of the gauge selection process, both under conditions of decentralized choices and under conditions of increased coordination.

A Spatial Model of Choice Among Techniques

W. Brian Arthur (1989) offers a non-spatial model of choice among techniques in the presence of two sorts of incentives: varying preferences for specific techniques and an interest in compatibility. He shows that, early in a selection process, the shares of different techniques may fluctuate randomly depending on the specific preferences of early adopters. If network benefits are unbounded, however, one technique eventually gains a sufficient lead in adoptions that the compatibility incentive overrides the preference that some new adopters have for the minority technique, causing the process to “lock in” to the majority technique.

Railways differ from the agents in Arthur’s model in that spatial relationships matter: railway lines have an incentive to use the same gauge as neighboring lines, not necessarily the majority gauge of the system as a whole. Furthermore, railway networks often fail to fulfill Arthur’s prediction of general standardization; uniformity often emerges at the regional but not national or continental level. One result of this regional standardization is that transition to larger-scale standardization requires the conversion of smaller common-gauge regions to the gauge of a larger region, a feature that need not be treated in Arthur’s model.

I therefore extend Arthur’s modeling framework substantially, both by specifying a spatial network within which agents are located and by providing for the possibility of conversion after initial choice of technique. I model individual railway lines as the cells of a lattice, define their mutually interdependent decision rules, and consider how the process evolves over time. Non-linearities in the model, related in part to conversion costs, make it analytically intractable, but I solve it using numerical simulation by computer. Technically, the model belongs to a class known as cellular automata, which have a long history in the modeling of nuclear chain reactions, ecological systems, and other phenomena in the natural sciences—as well as certain popular computer games. Economic applications, by contrast, have either been relatively abstract (Casti, 1989; Albin, 1998) or else focused on a limited range of issues such as land-use patterns and multiperson prisoners’ dilemmas.

This approach is in some respects similar to, if perhaps less elegant than, two other ways of modeling local (i.e., spatial) interactions, Markov random fields (or interacting particle systems) (David, 1993) and coordination games of learning by boundedly rational players (Ellison, 1993). In models using the former approach, agents switch their states (i.e., techniques) at random intervals to the states of randomly chosen neighbors. Such transition rules are not readily interpretable in terms of incentives for railway network integration, both because there is little justification for randomness in switching gauge and because these models involve no cost for switching. Such models also generally predict the emergence of a “continental” standard in every case. The latter approach, while it also captures the efforts of agents to coordinate their choices, is not generalizable to an appropriate network structure.

As Liebowitz and Margolis (1994, 1995) note, Arthur's modeling approach implicitly assumes substantial restrictions both on agents' foresight and on coordination or other means of internalizing the externalities among agents' choices. The same is true for the model here. Empirically, as we have seen, foresight and coordination were in fact severely restricted, particularly during the crucial early years of railways. Nevertheless, both factors became increasingly important over time. I therefore develop an extension of the model to account for these factors.

Assumptions

Simplifying assumptions for the model include the following: (1) available gauges are two in number ("broad" and "narrow") and remain the same throughout the process; (2) the network has a grid structure, rather than the empirically common "tree" structure with "trunk" and "branch" lines; (3) the network is located on a featureless plain, with no geographical barriers or concentrations of economic activity; (4) local lines are of equal size; and (5) parameter values are constant across both location and time. Section V discusses variations in parameter values that correspond to variations in the historical context.

The most problematic of these assumptions is probably that of the featureless plain. I adopt this assumption in order to be sure that results are not conditioned by a specific, arbitrary geographical context, but real-world geographical features have in fact often provided boundaries for common-gauge regions or defined routes over which traffic is heaviest and a common gauge is most valuable. Furthermore, narrow gauges have tended to be used more in mountains and in regions of low traffic demand. I return to this issue below.

The Network and Sequence of Events

The modeled continental network comprises 256 independent local railway lines, situated as the cells of a square lattice with 16 cells on each side (figure 1). This number of lines corresponds in order of magnitude to the number of original independent local railway companies in North America (the United States and Canada),[7] which was somewhat more than the number of independent agents in other major regions.

[Figure 1 here]

The modeled selection process takes place in two stages. First, local railways are constructed in random order, choosing their original gauges, until all sites in the lattice are filled. Empty sites next to existing lines are made ten times as likely to be filled as isolated sites, reflecting the tendency of potential routes that served as extensions of existing railways to be more valuable and more likely to be constructed. In the second stage of the process, local railways convert their gauges if their benefits from doing so exceed their costs. This division into stages is undertaken primarily for analytical purposes; still, in several historical cases, large regions were at least thinly spanned by railway networks before an appreciable number of conversions took place. In North America, for example, the first stage corresponds to the historical period from 1830 to 1864, the second stage 1865 to 1886 (Puffert, 2000).

Choice of Gauge by Local Railway Lines

Each local railway line chooses its gauge so as to maximize its value—that is, the present value of expected revenues minus costs. As in Arthur's (1989) model, this value is modeled as the sum of two terms, both of which depend on technique (gauge):

$$V(G) = D(G) + E(G),$$

where $\underline{G}$ represents gauge, $\underline{V(G)}$ is the line's value, $\underline{D(G)}$ is a technical-valuation function, and $\underline{E(G)}$ is a network-integration-benefit (or network-externality) function. Gauge $\underline{G}$ is chosen from the set $\{\underline{b},\underline{n}\}$, where $\underline{b}$ represents broad gauge and $\underline{n}$ represents narrow gauge.

For new railway lines, the technical function $\underline{D(G)}$ reflects the beliefs of each line's engineers and promoters about how gauge affects costs of construction, equipment, and operation, and how gauge affects quality of service and thus revenues. The function is modeled as a stochastically varying term, reflecting the variation in these beliefs. Following David (1987), I assume a continuous distribution of adopter types. Thus,

$$D(b) + \alpha$$

for broad gauge and

$$D(n) + \alpha + L,$$

for narrow gauge, where $\underline{\alpha}$ is a normalization term and $\underline{L}$ is a stochastic parameter that characterizes the extent to which the line's technical valuation of the narrow gauge exceeds that of the broad gauge. L is distributed uniformly over a unit interval that includes zero,

$$L \sim [\lambda – 1, \lambda],$$

so that 0 • λ • 1 represents the probability that any new railway line will prefer the narrow gauge. For the baseline simulation below, $\lambda = 0.5$, so that L falls in the range [–0.5,+0.5]. Some later simulations consider higher values for λ.

The network-integration-benefit function $E(G)$ represents the present value of the stream of incremental profits that a line expects to earn as a result of common-gauge connections with other lines. (These are the network externalities conveyed to the line by the other lines in its common-gauge network.) These profits are assumed to be proportional to traffic exchanges with other originating or destination railway lines, and traffic exchanges are assumed constant across each of two groups of other lines: direct neighbors and all other lines in the line’s common-gauge network. Thus, the incremental profits could be interpreted either as the total profits from each traffic-exchange connection, if breaks of gauge make traffic exchange prohibitively expensive, or otherwise as the savings in transshipment costs made possible by use of a common gauge. The function is expressed as

$$E(G) = \mu\, M(G) + \nu\, N(G),$$

where $M(G)$ is the number of neighboring lines with gauge G, $N(G)$ is the number of lines in the common-gauge network which the line would join by choosing G, and $\mu>0$ and $\nu>0$ are coefficients.

New railway lines without previously built neighbors form an expectation about future network integration benefits by considering any relatively nearby lines, specifically the established lines (and the sizes of their common-gauge networks) to which their future immediate neighbors will directly connect:

$$E(G) = \mu'\, M'(G) + \nu'\, N'(G),$$

where μ', ν', M', and N' are defined analogously to μ, ν, M, and N.

A new railway line chooses the broad gauge if and only if

$$V(b) > V(n).$$

A new railway line in an empty region chooses its gauge simply on the basis of its relative technical valuation. A line that connects to networks of both gauges evaluates the network integration benefits offered by each. A line with a higher technical valuation for one gauge (say, narrow) will choose the other (broad) if its difference in network integration benefits outweigh the difference in technical valuation. That is, the line chooses broad gauge if

$$E(b) - E(n) > D(n) - D(b) > 0$$

or $$\mu\, [M(b) - M(n)] + \nu\, [N(b) - N(n)] > L > 0.$$

Historically, new railway lines that connected to previously built railways have nearly always adopted the gauges of those railways. Thus, parameters for the model's baseline scenario give a greater value to the network benefits offered by even a small number of other lines than to the maximum difference in technical valuation. A scenario presented later considers a substantially greater relative difference in the technical valuation.

In the conversion phase of the process, railways switch their gauge, deterministically, if their potential gain in network integration benefits is greater than the cost of conversion:

$$E(a) - E(c) > C,$$

where $\underline{c}$ represents the "current" gauge (whether $\underline{b}$ or $\underline{n}$), $\underline{a}$ the "alternate" gauge ($\underline{n}$ or $\underline{b}$), and conversion cost $\underline{C}$ is assumed symmetric—the same in either direction. I no longer consider differences in technical valuation because, historically, original gauge preferences often became unimportant by the time of widespread conversion. In any case, maintaining the original preferences has no qualitative effect, and but little quantitative effect, on the model's results. Because a railway line's incentives for conversion may change as other lines convert, the conversion process continues until no further lines gain by converting.

The baseline set of results below are based on the following parameter values:

λ = 0.5, so that $\underline{L}$ is distributed uniformly on [–0.5, 0.5],

μ = 1.0,

ν = 0.08,

μ' = 0.1,

ν' = 0.02, and

C = 10.

In numerical simulation, values of $\underline{L}$ are supplied by a pseudo-random number generator that yields integers in the range $[-2^{15}, +(2^{15} - 1)]$; these integers are divided by 2^{16} to fall into the support of L. Zero is applied to the upper half of the distribution.

III. Results: A Sample Realization

Due to the model's stochastic features, each realization of the process develops differently. I present results in three stages: first using a sample realization that illustrates the model's mechanics and some principal features of potential outcomes, second using a Monte-Carlo experiment that shows large-sample results, and third with a series of Monte-Carlo experiments under varying parameter values. The latter experiments show both the robustness of qualitative results and the ways that quantitative results vary with variations in systematic causal factors, either across historical cases or over time within one case.

Figure 2 presents four "snapshot maps" of the developing process. The first three pioneering lines, in different parts of the lattice, choose gauges randomly, according to whether parameter L takes a positive or negative value (panel A). These lines then become the nuclei of expanding local common-gauge networks. The unlinked local lines in the "southwest" corner of panel A, labeled 4 and 5, are each close enough to the nearby networks that expectations of future links affect the relative values of different gauges. For example, the value of this expectation for line 4 is 0.26, which falls within the support of L and thus raises the ex-ante probability of choosing broad gauge from 0.5 to 0.76. Line 5 is affected by expectations of connection to both narrow-gauge network 2 and broad-gauge network 3.

[Figure 2 here]

Later (panel B), the original local networks both merge with other networks of the same gauge and also run up against networks of the other gauge. When all lines have been built (panel C), there is one large broad-gauge network, comprising 160 local lines, and two narrow-gauge networks, one with 15 and the other with 81 lines. During the conversion phase of the process (panel D), this difference in sizes, and hence in network integration benefits, outweighs the cost of conversion for each of the lines in the "northern" narrow-gauge network and for one line (marked "#") of the "southern" narrow-gauge network. The other narrow-gauge lines keeps their gauge. As a result, regional standard gauges emerge, but a continental standard does not.

The Impact of a Small Event

The modeled process is path dependent because slight variations in either the order of construction or the incentives of individual railways can lead to a large variation in the outcome. As it happens, this can be illustrated rather dramatically by a slight variation in the present realization in which the support of $\underline{L}$ is shifted to [–0.4999, +0.5001].[8] As a result, the realized value of $\underline{L}$ for one of the railway lines in figure 2, marked with an asterisk (*), becomes slightly positive rather than slightly negative, and the line chooses the narrow rather than the broad gauge. Neighboring railway lines, built later, then adopt the narrow gauge as well, forming a link between the northern and southern narrow-gauge networks rather than between the western and eastern broad-gauge networks (figure 3). The combined narrow-gauge network is then substantially larger than either of the broad-gauge networks, offering network integration benefits that eventually induce all the broad-gauge lines to convert their gauge. The process ends in standardization, and it does so at a different gauge than the majority gauge in the original realization.

[Figure 3 here]

Quantitative Evaluation of the Result

Using the first variation of the sample realization, let us note certain quantitative features of the result (table 2). Local lines make common-gauge connections with neighbors in 1,274 of 1,408 possible cases, which is to say that 90.5 percent of potential $\underline{M(G)}$ is realized. I call this a *local standardization index*. Furthermore, 56.9 percent of the potential value of $\underline{N(G)}$ is realized. Because this is the average proportion of other railway lines that each line has in its common-gauge network, I call this a *continental standardization index*.[9]

[Table 2 here]

An index of realized *network integration* benefits is effectively a weighted sum of the local and continental standardization indices. It takes a value of 64.0. Two adjustments must be made to this index to evaluate the relative economic efficiency of the outcome. First, the cost of 16 conversions is subtracted from realized network benefits. Second, the streams of benefits and costs are discounted on the assumption that the appropriate real interest rate is 4 percent (reasonable for relatively safe railway bonds in Britain and America in the nineteenth century) and that eight events (constructions or conversions) take place each "year." This adjustment reduces the final *index of ex-ante efficiency* to 59.3. Much of the reason that discounting reduces the value of the index is that, even if standardization does develop eventually, there are still costs in unrealized network integration benefits in the short run.

Finally, an *index of ex-post efficiency* compares realized network integration benefits with the benefits—net of further conversion costs—that could be realized by converting the 80 remaining narrow-gauge lines to the broad gauge; the sunk costs of earlier conversions do not enter the calculation. The index takes a value of 72.8, showing the outcome is inefficient from an *ex post* as well as *ex ante* point of view. This inefficiency is an indication of possible unrealized profit opportunities that, depending on transaction costs, may be available to an agent that internalizes the mutual externalities among railways (Liebowitz and Margolis, 1994, 1995). I consider the effects of this on the model's results in section VI.

IV. Results of the Model: A Monte Carlo Experiment

This sample realization of the gauge selection process shows one possible way that the process may evolve. In order to investigate the range of possible outcomes, a Monte Carlo experiment was conducted: The process was repeated for 1,600 realizations, using the same values of the model's parameters, but with stochastic variation in both the order of construction and the preferred gauge of each local railway line.[10]

As the first column of table 3 shows, on average, about half the lines have adopted each of the two gauges by the end of both the construction phase of the process (50.9 percent narrow gauge) and the conversion phase (51.6 percent narrow gauge). Confidence intervals for these results, which indicate the range of values within which the model's "population mean" is likely to fall, both include the value 50 percent, as expected given the symmetric position of the gauges in the model. However, the sample standard deviations of these mean results are quite large (27.8 and 44.9 percent, respectively), indicating that most individual realizations generated a substantial majority of one or the other gauge. This distribution of results is broad and bi-modal (figure 4), showing that the process is *symmetry-breaking*; it nearly always "tips" to favor one gauge or the other. By the end of the conversion phase, most of the weight of the distribution lies at the two extremes.

[Table 3 here] / [Figure 4 here]

This result differs markedly from the narrow, uni-modal, asymptotically Gaussian distribution that would result if each local railway line chose its gauge simply according to its technical valuation, without reference to network integration benefits. In that case, according to statistical theory, the variance of the mean proportion of narrow gauge in each realization is $(.5)^2/256$ or 0.00098, yielding a standard deviation of 0.031 or 3.1 percent. In the experiment, the estimated variance and standard deviation for the construction phase are, respectively, 80 and 9 times these theoretical values. As 80 is far beyond the (modified) chi-square critical value of 1.1 (1-percent significance level), one can quite conclusively reject the hypothesis that the result of the process is indistinguishable from that of random choices.

Stochastic events, in both the order of construction and the gauges favored by local railway lines, make the process *path-dependent*, and the result is *unpredictable* at the outset for a hypothetical observer who knows only the general distribution of preferences and the structure of the process. (Granted, such an observer would have a knowledge of the process greater than that of the agents.)

By the conclusion of the conversion phase, the process results in a standard gauge in 78.8 percent of the realizations. The index of continental standardization indicates that, taking these realizations together with those not resulting in an overall standard, local railway lines end up in a common-gauge network together with an average of 90.0 percent of all other lines. The index of local standardization shows that local lines share a common gauge with 97.5 percent of their immediate neighbors. Thus, even realizations that do not end up with a continental standard still exhibit a high degree of *local standardization*.

Failure to generate a global standard means that there are unrealized potential network integration benefits and thus *inefficiency* in the outcome relative to other outcomes that were available for the process ex ante. In this experiment, an average of 91.6 percent of potential network integration benefits are realized by the end of the process. Furthermore, nearly all the realizations that eventually result in standardization nevertheless have both short-term diversity and conversion costs. Taking these additional sources of inefficiency into account, an average of 76.6 percent of the potential benefits of network integration are realized.

Perhaps as notable as these average figures is the dispersion of results, as indicated by the sample standard deviations. The fact that several of these statistics lie within about one sample standard deviation of 100 indicates that the lower tails of the distributions have substantial weight.

Several qualitative results of the model—symmetry-breaking, path dependence, unpredictability, and potential inefficiency—correspond to results of Arthur’s non-spatial model (Arthur 1989). An important difference in the spatial case is the possibility of local standardization together with continuing continent-level diversity.

V. Variations of the Model, with Applications to History

A series of further Monte Carlo experiments both confirms the robustness of qualitative results and explores how the distribution of quantitative results depends on variations in the model’s parameter values and structure. These variations correspond to variations in the historical context. Experiments also yield further insight into the effects of particular contingent events, especially early choices of gauge.

Impact of Early Events

A matter of particular interest for the interpretation of history is the impact of early choices on the subsequent development of the process. As in Arthur’s (1989) non-spatial model, early events have a disproportionate effect (table 4). The gauge that happens to be chosen by the first line built tends, on average, to be adopted by nearly two-thirds (66.3 percent) of all lines built thereafter. Furthermore, more than two-thirds of the trials that result in standardization (56.7 out of 81.0 percent of all trials) do so using the gauge of the first line.

[Table 4 here]

A related question is the impact of intentional—or even accidental—coordination among early agents in different regions. In cases where the second new railway line adopts the same gauge as the first, more than three-fourths of all lines (78.7 percent) eventually adopt that gauge and virtually 70 percent of all trials result in standardization at that gauge. The overall probability of standardization also is significantly greater than under the baseline scenario, and network integration and efficiency indices are greater as well. These results obtain still more strongly for cases in which the first four lines adopt the same gauge.

The impact of early gauge choices depends on the line's location within the lattice. When the first line is built in the center of the lattice, rather than in a random location, both the mean proportion of lines using that gauge and the likelihood of that gauge becoming the standard are significantly greater—76.3 and 66.8 percent, respectively. Interestingly, however, the estimated overall probability of a standard emerging is not significantly greater than in the baseline scenario, and neither are the indices of standardization and network integration. Finally, four lines built to the same gauge in the corners of the lattice have much less effect on the process than four lines built in random locations.

Together, this first group of experiments supports the interpretation that the dominance of the Stephenson gauge in Britain, continental Europe, North America, and elsewhere is due to its adoption early and in several parts of those regions, not due to an inherent superiority. The gauge's adoption in more central locations—clearer in Britain and the Continent than in North America—also contributed to its success.

Impact of Variations in Parameters

A further series of experiments addresses the effects of variations in the model's parameters related to the relative technical valuation of gauges, network integration benefits, and conversion cost—all matters that varied among historical contexts (table 5). First, suppose, without loss of generality, that the narrow gauge is valued more highly by a majority of railway lines. Does that gauge necessarily predominate in the outcome? When the narrow gauge is preferred by 62.5 of the population of potential adopters, it ends up being adopted by 78 percent; when preferred by 75 percent, it is adopted by 90 percent. Thus, in addition to being symmetry-breaking, the process is "asymmetry-enhancing." The gauge preferred more often is likely to gain an early lead in adoptions, offering network integration benefits that induce later adopters to choose that gauge even if they have a greater technical valuation for the alternative gauge.

[Table 5 here]

Also of interest, the probability of attaining a standard increases given asymmetric preferences, and so do the measures of standardization, network integration, and efficiency. Nevertheless, it remains possible for the process to standardize on the less often preferred gauge, as happened in 5.2 percent of trials even when 75 percent preferred narrow-gauge. For the interpretation of any specific historical case, this means that we cannot infer much about the initial distribution of relative valuations simply from the end result of the process.

Next, historically, relative preferences for broader and narrower gauges changed over the course of selection processes. How flexible could these processes have been to those changes? We consider a shift in valuation such that, after either 64 or 72 lines have all been built at the broad gauge (without loss of generality), all new lines prefer narrow gauge by a relative valuation equivalent to the network integration benefits offered by five common-gauge neighbors (or one neighbor plus a 50-line common-gauge network). Results show that a shift in valuations after 64 lines have been built leads to the widespread adoption of the narrow gauge, but a shift after 72 lines have been built does not. As in Arthur's nonspatial model, the process becomes inflexible to the change in incentives—it "locks in" to the first gauge introduced. Historically, newly preferred gauges have been able to get a foothold only where previous railways are sparse. Efforts, like that of Britain's Great Western Railway, to introduce new gauges after other gauges have gained a substantial lead have always failed.

Perhaps the greatest problem in calibrating the model to historical cases is to judge the relative strength of the valuation of specific gauges against network integration benefits. Moreover, this relative valuation varied historically, as some railway builders had strong preferences for broad or narrow gauges. A fifth experiment gives the modeled valuation of specific gauges a ten-times greater weight. Perhaps surprisingly, most results are statistically indistinguishable from baseline results. Only the indices of local standardization and ex-ante efficiency take statistically significant lower values. The model's results are robust both to uncertain assumptions and to a range of differences in historical cases.

Next, in numerous historical cases relative gauge preference was arguably endogenous, in part because apprentice engineers that gained experience with a gauge in building one line continued to use it when they became chief engineers for later new lines. I make parameter $\underline{\lambda}$

an endogenous function of the number of previously built railways of each gauge: $\underline{\lambda}$ takes the value $^{(N_n+2)}/(N_T+4)$, where $\underline{N_n}$ is the number of previously built narrow-gauge lines and $\underline{N_T}$ is the number of previously built lines of both gauges. The process begins with a 50-percent probability that each gauge is preferred, but as it proceeds, these probabilities approach each gauge's share of established railways. As a result, relative technical valuation itself has positive feedbacks, and the process "tips" more quickly to favor one gauge or the other. More of the realizations result in standardization, and all indices take higher values.

The form and level of network integration benefits also affects the model's results (panel B of table 5), but this is less of interest for understanding variations in historical context than for understanding the dynamics of the model. First, if a railway line's network integration benefits depend simply on the size of its common-gauge network, with no additional benefit resulting from neighboring lines, then lines form significantly fewer common-gauge connections with neighbors. Eliminating benefits resulting from the size of the network greatly reduces the incidence of standardization and all related indices. Doubling these benefits naturally increases all these statistics. Finally, elimination of the expectations component of gauge choice yields results indistinguishable from those of the baseline scenario.

The level of conversion costs relative to network integration benefits has a substantial effect on the likelihood that early diversity is resolved (panel C of table 5). With zero conversion costs, all realizations result in standardization. As costs increase to 50, 100, and 150 percent of the level in the baseline scenario, the likelihood of standardization and all indices decline. One factor that favored the rapid resolution of diversity in the United States was the easy convertability of its track, where rails were usually spiked directly to wooden cross ties and could readily be moved laterally. On Britain's GWR system, by contrast, most rails were laid on longitudinal sleepers buried in the ground, and it was much more costly and disruptive to service to change the gauge.

Effects of Variations in the Structure of the Model

A final series of experiments considers several variations in the structure of the model (table 6). First, an experiment establishes a standard of comparison for the baseline results by showing the extent to which network integration emerges even when not intentionally pursued—that is, when railway lines choose their gauges on the basis of their technical valuations alone, that is, randomly. Given the choice of only two gauges, substantial common-gauge regions must emerge for simple reasons of topology.[11] An average of 14 common-gauge networks form, and the mean index of global standardization shows that, on average, lines find themselves in a common-gauge network with 36.6 percent of other lines in the lattice. Clearly a few of the common-gauge networks are relatively large.

[Table 6 here]

A related experiment considers the extent to which incentives for gauge conversion are sufficient to resolve the diversity that results from random initial choices. These incentives lead to eventual standardization in over 30 percent of trials and reduce the average number of common-gauge networks from 14 to two.

An experiment that overlaps the construction and conversion phase of the process yields a statistically insignificant difference in nearly all results. Next, a scenario with a less connected network structure, in which each line has four rather than six neighbors, leads to substantially less standardization, integration, and efficiency. Finally, modeling a smaller lattice size—12 by 12 for 144 lines—also generates lower values for these statistics. Smaller and less densely connected networks prove less likely to result in a standard.

Modeling and Historical Realism

While these variations in the model's parameters and structure account for some variations in the historical context, no version of the model corresponds closely to any specific historical case or geographic setting. The most obvious omissions are those of cities and other concentrations of economic activity, major trunk routes, and physical and political geographic features that divide a continent into subregions and sometimes, particularly in the case of mountains, encourage use of particular gauges. These features have often helped define regions that adopted common gauges either initially or through conversion. Most importantly, standardization of gauge on major interregional trunk routes has usually been the first step in converting variant-gauge regions, including in Britain, in most North American cases, in the Netherlands, and recently in Australia.

Nevertheless, a more general modeling approach assures that results depend not on specific narrow assumptions but rather on the general logic of positive feedbacks within a spatial network. Application of the results to interpretation of specific historical episodes requires attention to specific geographic and other details, but there is no reason why the essential logic of the gauge selection process should be affected. Further research on the effects of specific geographic features could perhaps be most useful in helping to specify plausible counterfactuals for specific historical episodes.

The model’s assumption of only two gauges applies well enough to places such as Britain and Continental Europe, where Stephenson-gauge networks were interrupted by separated pockets of other gauges. But it does not apply well to North America, where three different gauges came together in some places. If North American railways had had only two gauges to choose from, then much less diversity would have developed, as more local common-gauge regions would have had to merge into other regions of the same gauge—an implication of the four-color theorem in map-making and graph theory. Nevertheless, this feature of the model does not affect the essential process of emergence and resolution of diversity.

VI. Extensions to the Model: Foresight and Coordination

More important for the interpretation of history, how are the model’s results affected when agents exercise greater foresight and coordination? First, suppose that all agents know from the beginning the future benefits of long-distance network integration and thus the value of a common gauge. In that case, a simple game-theoretic framework shows that agents will standardize from the beginning; strategic, preemptive commitment determines which gauge becomes the standard. Alternatively, if new railway lines in empty regions at least take account of the likelihood that distant common-gauge networks will eventually expand into their regions, then they will often adopt the gauges of those networks, and less diversity will develop. Historically, railway builders often undervalued future network integration, but in many cases they did, in fact, choose gauges in anticipation of future connections.

Results are also affected when agents internalize their mutual externalities, influencing each other's choices through side payments, through overlapping ownership, or simply through coordinating their decisions—all quite relevant historically, particularly during the conversion phase of the process. Still, although externality internalization may greatly enhance the resolution of diversity, most results of the baseline model still hold.

Consider, first, the net increase in network integration benefits that a railway line, situated on the edge of the smaller of two common-gauge networks, can realize for itself through conversion. This increase is proportional to the difference in sizes of the networks, $\underline{N(a)}$–$\underline{N(c)}$ in the notation introduced earlier, and is depicted as the "decentralized choices" function in figure 5. It is assumed for simplicity that the line has three neighbors of each gauge.

[Figure 5 here]

This line's conversion increases not only its own network integration benefits but also, it can readily be shown, an equal net level of benefits externally for other lines, subtracting losses to lines in the line's former network from gains to lines in the new one. If the affected lines offer side payments reflecting their gains (or potential losses), then the maximum coverable conversion cost is doubled (figure 5, "side payments").

Both of these conversion schemes reflect only *marginal* effects—the gains from converting one line only. If the entire minority-gauge network can be converted, then it is relevant to compare the *average* gain per line to the cost (per line) of conversion. Accounting for all external effects, the average social value of converting can be shown to be as indicated by the "encompassing coalition" function—so named because it indicates the maximum contribution that all railway lines together could make to the costs of conversion.[12] Considering only the benefits realized by the converting lines (not lines already using the majority gauge), the average value is given by the "minority coalition" function. The area between the minority-coalition and decentralized-choices functions reflects situations where all members of this coalition are better off if all convert, although none gains by converting alone. By converting together, they gain the benefits of compatibility with the larger network without losing the benefits of compatibility among themselves. As it happens, any of these internalization schemes would suffice to resolve the diversity remaining at the end of the sample realization discussed above (figure 2, panel D), indicated in figure 5 with an asterisk (*).

How do these internalization schemes affect the model's results? In each case, the function showing benefits to conversion, and thus maximum coverable conversion costs, increases with the difference in network sizes, $\underline{N(a)} - \underline{N(c)}$. If, on the one hand, unit conversion costs are high enough relative to potential gains in network integration benefits (at levels above 20.4 in figure 5, if encompassing coalitions are feasible), then whether diversity is resolved still depends on the degree of asymmetry in original gauge adoption. The qualitative results of the baseline model continue to hold.

On the other hand, two of the functions for externality-internalization schemes have positive intercepts, unlike the decentralized-choices function. This means that, if unit conversion costs are low enough relative to potential gains in network integration benefits, then externality internalization makes it possible to resolve any degree of early diversity. Fifty-fifty splits, which are not resolved by lines acting alone, can be resolved by converting groups of lines together. This last conclusion does contrast with results of the baseline model.

Historically, nothing resembling an encompassing coalition has ever formed for the purpose of resolving differences in gauges. Nor have more than a few connecting lines of the majority gauge ever contributed side payments for the conversion of minority-gauge routes. Transactions costs would naturally be high in organizing larger schemes, as suggested by the long history of failures to resolve Australia's diversity. "Minority coalitions" have formed, however, most notably in converting the railways of the southeastern United States in 1886, and also in several cases involving small groups of railways.

By far the most important way that externalities have been internalized has been through common ownership—the formation of interregional, initially multi-gauge railway systems. Britain's diversity was resolved after the Great Western Railway system incorporated numerous Stephenson-gauge routes. In North America, important conversions were undertaken by interregional trunkline systems such as the Pennsylvania Railroad and the Illinois Central Railroad. Conversion of Australia's broad-gauge railways followed takeover of the railways of South Australia by the Commonwealth (national) Railways, owner of the

Stephenson-gauge transcontinental line.

In contrast to the model and its extension, with their assumption of a featureless plain and undifferentiated traffic demand, these historical interregional systems have comprised ad-hoc groupings of routes and regions with particular concentrations of traffic. Nevertheless, the essential conclusions of the model regarding the relation between conversion costs and potential network integration benefits should still hold.

As the example of Australia—and the lack of positive examples elsewhere—shows, government-owned railway systems may have difficulty internalizing their mutual externalities through side payments or (international or interstate) takeovers. The lack of internalization mechanisms may hinder conversion that would be worth the cost—perhaps someday in Spain. Or, new internalization mechanisms may be developed, perhaps in Spain's case within the framework of the European Union.

To conclude, externality internalization can help assure that the resolution of early, path-dependent diversity takes place where the potential gains in efficiency are greatest. In North America, cooperation and system-building led to a rapid conversion as demand grew for interregional transport. Indeed, given that this happened relatively early in the development of both traffic demand and interregional systems, one may conclude that even a much greater diversity of gauge, had it happened to develop, would eventually have been resolved. In other historical cases, where either the benefits of standardization or opportunities for internalization have been less, early path-dependent diversity has persisted, at some cost in efficiency relative to what standardization from the beginning would have yielded. One may presume that diversity is resolved whenever the cost of its persistence exceeds the cost of remediation—including transactions cost in organizing the internalization of externalities.[13]

VII. Conclusion

Historically as well as in the model, original regional gauge choices were drawn essentially as random samples from a range of available practices, and the benefits of compatibility led subsequent, connecting lines to adopt the same gauges. As a result of these positive feedbacks, common-gauge regions of various sizes emerged. The resolution of diversity among these regions has depended both on the extent of early diversity and on the relation between potential network integration benefits and cost of conversion. Thus the selection of regional standard railway gauges has been path dependent, both in which gauges emerged as standards and in the extent of diversity that emerged and persisted.

The gauge now used on nearly 60 percent of the world's railways, like other gauges, was not primarily the result of fundamental incentives, systematic optimization, or a market test but rather of a series of contingent events—even of historical accidents—reinforced by positive feedbacks. The relative merits of different gauges have, of course, been tested by experience, but not in a way that has selected the best as a regional standard, largely because the costs of conversion have been greater than the potential gains. Experience has, however, several times refuted expectations that new variant gauges would offer technical advantages outweighing the costs of diversity. Experience has also shown broader gauges to be generally better than narrower, causing regret in regions where particularly narrow gauges emerged as standards.

More often, experience has caused regret over the emergence of diversity, which has generated costs first of coping with breaks of gauge and then, sometimes, of converting whole regions. The resolution of diversity through conversion was, of course, a matter of systematic optimization, and it often happened through the sort of coordinating, externality-internalizing behavior expected by Liebowitz and Margolis (1994, 1995). These authors' view that path dependence might depend on the lack of early foresight—here, foresight into the later importance of long-distance, large-scale network integration—also receives empirical support.

The case of track gauge also supports both Arthur's (1989, 1994) general modeling approach and his proposition that path-dependent processes can yield inefficient outcomes. In contrast to the results of Arthur's non-spatial models, however, the case offers two lessons for the emergence of standards in spatial networks. First, regional standards emerge, but "global" or continental standards do not necessarily do so unless some regions are converted ex-post. Second, the potential inefficiency of a spatial path-dependent process may lie much more in the persistence of diversity than in selection of a suboptimal technique.

Both of these lessons also apply to other technical features of railways. For example, trains that pass through the Channel Tunnel between London and Paris or Brussels have had to cope with three different electrical power systems (varying in voltage, alternating or direct current, AC frequency, and collection mechanism), five different train-control and signaling systems, and differences in loading gauge (clearance dimensions) and other parameters. As a result, duplicate technical systems have raised costs, and train performance could not be optimized for any part of the system. Similar variations hinder the development of high-speed train service elsewhere within the Stephenson-gauge region of Europe, but the growing importance of these variations is leading to their partial resolution (Puffert, 1993, 1994).

More broadly, these lessons apply to other spatial networks—such as for transportation, communication, and electrical power distribution—as well as to networks with non-spatial graphical structures (patterns of connectedness), but in which each agent has direct network interactions with a relatively small subset of other agents. This arguably includes most empirical networks, including the "virtual" networks often considered in discussions of network externalities (Katz and Shapiro, 1994; Economides, 1996).

In view of recent disputes over path dependence (Liebowitz and Margolis, 1994, 1995), it is worth noting that the inefficiency discussed here is primarily the result not of market or institutional failure but rather of early lack of foresight combined with positive feedbacks that lend increasing impact to early agents' choices. As a result, different possible sequences of contingent events would yield outcomes differing in their relative efficiency, and the process has little tendency to converge to its optimal potential outcome. The upper bound cost of potential inefficiency is the cost, including transaction costs, of full remediation.

Whether market failure, narrowly defined in terms of a difference between the (foreseeable) private and social costs—and benefits—of an agent's actions, also played a substantial role is a matter for future empirical investigation. It is notable that side payments, appropriation, and cooperation internalized the external effects of railway lines' choices of gauge sufficiently to yield socially optimal conversions of gauge in numerous cases. One may presume, however, that transactions costs have hindered such actions in other cases. Nevertheless, it is far from clear that a public agency would often possess enough information to improve on the cooperative actions of railway operators.

It is noteworthy, moreover, that the clearest example of institutional failure was not one of markets but of governments—specifically of the separate Australian colonies (later states) and the British colonial administration. If private firms do not necessarily internalize their mutual externalities optimally, then separate states may be even less likely to do so, being less attentive to market incentives and rarely subject to takeover by an interregional system.

Notes

[1] EH.RES list archives, http://www.eh.net/, October 1996 through April 2001. Admittedly, this count reflects in part the particular interests of a relatively small number of vigorous participants.

[2] Some of these issues arise implicitly, but are not directly examined, in case studies of path-dependent selection among alternative techniques in nuclear power (Cowan, 1990), electrical power distribution (David, 1990), videocassette recording (Cusumano et al., 1994), and pest control (Cowan and Gunby, 1996). See also Scott's (2001) discussion of Britain's "coal car problem."

[3] For example, the president of the (U.S.) Burlington Northern Railroad wrote in 1978 that "if we had it to do all over again we'd probably build them with the rails at least 6 feet apart," although another authority wrote at the time that, although broader gauges are sometimes advantageous, for general service the Stephenson gauge is not far from optimal (Hilton, 1990, p. 37). Engineers whom I interviewed at the Association of American Railroads and American Railway Engineering Association also favored broader gauges.

[4] For a more detailed account see Puffert (1991). For events in North America, see Puffert (2000).

[5] The PRR actually converted its eastern trunk route from 4'8.5" to 4'9" and its western routes from 4'10" to 4'9.5", changing a problematic 1.5" difference in gauge to a series of manageable half-inch differences. The PRR reduced the gauge of its western routes to 4'9" during the late 1870s after most of the independent Ohio railways had reduced their gauge by at least half an inch. The 4'9" gauge remained in use until after 1900. The Lake Shore and Michigan Southern Railroad also played a role in reducing Ohio's gauge (Puffert, 2000).

[6] Economies of scale in rolling stock for particular gauges are exhausted at low levels, particularly for rolling stock other than locomotives, which often have differed among gauges only in their wheel trucks.

[7] The correspondence is approximate. The 1880 U.S. Census (Shuman, 1883) lists 1,174 individual railway companies, and Canada had several dozen more. Many of these companies were only short extensions of a single other railway, and others were owned from the start by other railways. Thus these did not independently affect the dynamics of the process.

[8] The process is numerically simulated using the same sequence of pseudo-random numbers in both cases.

[9] The calculation is available from the author. Intuitively, each railway line's contribution to this index is the proportion of other lines in its common-gauge network. The index as a whole averages these contributions.

[10] The realizations differ in the series of pseudo-random numbers used in the numerical simulation.

[11] This is an implication of graph theory's four-color theorem, which holds that up to four colors are needed to color an arbitrary two-dimensional map in such a way that no adjoining regions have the same color.

[12] For simplicity, the effect of eliminating breaks of gauge among immediate neighbors is here neglected.

[13] This proposition, derived from Liebowitz and Margolis (1994, 1995), is admittedly non-testable and tautological, as any failure to remedy can be ascribed to transaction costs.

References

Albin, P.S. (1998). *Barriers and Bounds to Rationality: Essays on Economic Complexity and Dynamics in Interactive Systems*. Princeton: Princeton University Press.

Arthur, W.B. (1989). Competing Technologies, Increasing Returns, and Lock-in by Historical Events. *Economic Journal*, **99**, 116-31.

—— (1994). *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.

Carlson, R.E. (1969). *The Liverpool and Manchester Railway Project 1821-1831*. New York: A.M. Kelley.

Casti, J.L. (1989). *Alternate Realities: Mathematical Models of Nature and Man*. New York: Wiley.

Cowan, Robin (1990). Nuclear Power Reactors: A Study in Technological Lock-in. *Journal of Economic History*, **50**, 541-67.

Cowan, Robin and Philip Gunby (1996). Sprayed to Death: Path Dependence, Lock-in and Pest Control Strategies. *Economic Journal*, **106**, 521-42.

Cusumano, Michael A., Yiorgos Mylonadis, and Richard S. Rosenbloom (1992). Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta. *Business History Review*, **66**, 51-94.

David, Paul A. (1985). Clio and the Economics of QWERTY. *American Economic Review*, **75** (Papers and Proceedings), 332-37.

—— (1987). Some New Standards for the Economics of Standardization in the Information Age. In P. Dasgupta and P. Stoneman (eds.) *Economic Policy and Technological Performance*. Cambridge: Cambridge University Press.

—— (1990). Heroes, Herds, and Hysteresis in Technological History: Thomas Edison and the Battle of the Systems Reconsidered. *Journal of Industrial and Corporate Change*, **1**, 129-180.

—— (1993). Path-dependence and Predictability in Dynamic Systems with Local Network Externalities: A Paradigm for Historical Economics. In D. Foray and C. Freeman (eds.) *Technology and the Wealth of Nations: The Dynamics of Constructed Advantage*. London: Pinter.

Economides, N. (1996). The Economics of Networks. *International Journal of Industrial Organization*, **14**, 673-99.

Ellison, G. (1993). Learning, Local Interaction, and Coordination. *Econometrica* **61**,1047-71.

Farrell, J. and G. Saloner (1985). Standardization, Compatibility, and Innovation. *Rand Journal of Economics*, **16**, 70-83.

Great Britain, Gauge Commission (1846). *Report of the Gauge Commissioners*. London: T.R. Harrison.

Harding, E. (1958). *Uniform Railway Gauge*. Melbourne: Lothian.

Haywood, R.M. (1969). The Question of a Standard Gauge for Russian Railways, 1836-1860. *Slavic Review*, **28**, 72-80.

Hilton, George W. *American Narrow-Gauge Railroads*. Stanford: Stanford University Press, 1990.

Katz, M.L. and Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. *American Economic Review*, **75**, 424-40.

—— and —— (1994). Systems Competition and Network Effects. *Journal of Economic Perspectives*, **8**, 93-115.

Liebowitz, S.J. and Margolis, S.E. (1990). The Fable of the Keys. *Journal of Law and Economics*, **33**, 1-25.

—— and —— (1994). Network Externality: An Uncommon Tragedy. *Journal of Economic Perspectives*, **8**, 133-50.

—— and —— (1995). Path Dependence, Lock-In, and History. *Journal of Law, Economics, and Organization*, **11**, 205-26.

Puffert, D.J. (1991). The Economics of Spatial Network Externalities and the Dynamics of Railway Gauge Standardization. Ph.D. dissertation, Stanford University.

—— (1993). Technical Diversity and the Integration of the European High-Speed Train Network. In J. Whitelegg, S. Hultén, and T. Flink (eds.) *High-Speed Trains: Fast Tracks to the Future*. Hawes, North Yorkshire: Leading Edge.

—— (1994). The Technical Integration of the European Railway Network. In A. Carreras, A. Giuntini, and M. Merger (eds.) *European Networks, 19th-20th Centuries: Approaches to the Formation of a Transnational Transport and Communications System*. (Proceedings, Eleventh International Economic History Congress). Milan, Italy: Universita Bocconi.

—— (2000). The Standardization of Track Gauge on North American Railways, 1830-1890. *Journal of Economic History* **60**, 933-60.

Scott, P. (2001). Path Dependence and Britain's "Coal Wagon Problem." *Explorations in Economic History* **38**, 366-85.

Shuman, A.E. (1883). Statistical Report of the Railroads in the United States. In *Report on the Tenth Census,* vol. 4: *Report on the Agencies of Transportation in the United States*. Washington, D.C.: United States Department of the Interior, Census Office.

Smiles, S. (1868). *The Life of George Stephenson and of his Son Robert Stephenson*. New York: Harper.

Table 1. Principal Railway Track Gauges, 2000

Gauge			Proportion of
English	Metric	Major countries and regions	world total[1]
(ft.-in.)	(mm.)		(percent)
2’6”	762	China*[2], India*	1.7
3’0”	914	Colombia, Guatemala, Ireland*	0.6
3’3.37”	1000	East Africa, Southeast Asia*, Argentina*, Brazil*, Chile*, India*, Pakistan*, Spain*, Switzerland*	8.8
3’6”	1067	Southern Africa, Southeast Asia*, North Africa & Middle East*[3], Australia*, Japan*, New Zealand, Newfoundland	9.0
4’8.5”	1435	Europe*, North America, North Africa & Middle East*, Argentina*, Australia*, Chile*, China*, Japan*	58.2
5’0”	1524	Former USSR, Finland, Mongolia	12.8
5’3”	1600	Australia*, Brazil*, Ireland*	1.2
5’6”	1676	Argentina*, Chile*, India*, Pakistan*, Portugal & Spain*[4]	7.0

Notes: *Countries or regions with more than one gauge. [1]Percentages add to less than 100 due to additional, rare gauges. [2]750 mm. [3]1055 mm. [4]Originally 1672 mm.; now 1668 mm.

Sources: *Jane’s World Railways*; *Railway Directory and Yearbook*

Table 2
Sample Realization: Quantitative Characteristics at End of the Process

Characteristic	Value
	Scale of 100
Network characteristics:	
Local standardization index (realized sum, numbers of common-gauge neighbors, $\Sigma_i M_i(G)$)	90.5
Continental standardization index (realized sum, sizes of common-gauge networks, $\Sigma_i N_i(G)$)	56.9
Economic characteristics (network integration and efficiency):	
Network integration index (realized integration benefits, $\Sigma_i \, [\mu M_i(G) + \nu N_i(G)]$)	64.0
Preliminary ex-ante efficiency index (the above minus conversion cost expended, $\Sigma_i \, [\mu M_i(G) + \nu N_i(G)] - \Sigma_i C_i$)	61.6
Ex-ante efficiency index (the above with discounting of benefits and costs)	59.3
Ex-post efficiency index (realized $\Sigma_i \, [\mu M_i(G) + \nu N_i(G)]$, relative to potential value minus cost of additional needed conversions)	72.8

Note: Subscripts i index local railway lines.

Table 3
Numerical Simulation: Summary Results of Monte Carlo Experiment

	Mean proportion narrow gauge	Proportion of trials resulting in standardization	Mean number common-gauge networks	Mean indices of network characteristics			
				Continental standardization	Local standardization	Network integration benefits	Ex-ante efficiency
	——— *percent* ———			—————— *scale of 100* ——————			
Construction Phase							
Estimate:	50.9	0.2	3.73	63.2	86.4	68.1	67.0
Sample standard deviation:	27.8	5.0	1.18	13.2	4.4	11.0	10.8
95% confidence interval (±):	1.4	0.2	0.06	0.6	0.2	0.5	0.5
Conversion Phase							
Estimate:	51.6	78.8	1.26	90.0	97.5	91.6	76.6
Sample standard deviation:	44.9	40.9	0.55	19.4	5.0	16.3	11.5
95% confidence interval (±):	2.2	2.0	0.03	0.9	0.2	0.8	0.6

Notes: Number of trials = 1,600. The “95% confidence interval” is 1.962 (the relevant critical value of t for a two-sided test) times the standard error of the estimate.

Table 4
Monte-Carlo Experiments: The Impact of Early Choices of Gauge

(Point estimates for results at the end of the process)

	Mean	Proportion of		Mean	Mean indices of network characteristics			
	propor-	trials resulting in		number of		Conti-	Network	
	tion of	standardization		common-	Local	nental	integra-	*Ex-ante*
	first	First	Either	gauge	standard-	standard-	tion	effici-
Experimental scenario	gauge	gauge	gauge	networks	ization	ization	benefits	ency
		Percent				*Scale of 100*		
First gauge adopted by ...								
One line	66.3*	56.7*	81.0	1.25	97.8	91.0	92.4	77.6
First two lines	78.7*	69.8*	83.5*	1.21	98.1*	92.1*	93.4*	79.0*
First four lines	90.7*	86.7*	92.8*	1.09*	99.1*	96.5*	97.1*	84.2*
First line (in center)	76.3*	66.8*	81.7	1.23	97.8	91.3	92.7	77.6
First four lines (in corners)	66.0*	57.2*	83.3*	1.26	97.8	91.9	93.2	77.2

*Results significantly different from baseline result at 5-percent level.

Note: The baseline experiment used 1,600 realizations of the process; others each used approximately 600 realizations.

Table 5

Monte-Carlo experiments: Variations in technical and network parameters

(Point estimates for results at the end of the process)

	Mean	Proportion of		Mean	Mean indices of network characteristics			
	propor-	trials resulting in		number of		Conti-	Network	
	tion of	standardization		common-	Local	nental	integra-	*Ex-ante*
	narrow	Narrow	Either	gauge	standard-	standard-	tion	effici-
Experimental scenario	gauge	gauge	gauge	networks	ization	ization	benefits	ency
	——— *Percent* ———				———*Scale of 100*——— —			
Baseline scenario	*51.6*	*40.9*	*78.8*	*1.26*	*97.5*	*90.0*	*91.6*	*76.6*
A. Gauge preference:								
Narrow preferred by 62.5%	77.9*	69.3*	83.8*	1.20*	98.1*	92.3*	93.5*	80.2*
Narrow preferred by 75%	90.5*	85.0*	90.2*	1.12*	98.9*	95.3*	96.0*	86.2*
Shift in preference to narrow ...								
after first 64 lines	57.4*	52.5*	87.7*	1.27	98.1*	94.3*	95.1*	80.0*
after first 72 lines	27.9*	23.3*	87.5*	1.23	98.1*	94.5*	95.3*	82.0*
Ten-times stronger valuation	51.9	40.3	77.8	1.28	94.8*	89.2	90.4	65.4*
Endogenous preference	51.7	47.8*	92.0*	1.10*	99.0*	96.2*	96.8*	87.7*
B. Network benefits:								
No neighbor benefit	48.8	37.3	76.0	1.69*	95.7*	88.9	88.9*	66.7*
No network-size benefit	50.3	0.5*	1.0*	2.95*	89.8*	54.9*	89.8*	89.9*
Doubled network-size benefit	50.5	48.8*	96.3*	1.04*	99.6*	98.2*	98.4*	79.4*
No expectations	49.7	39.3	79.7	1.27	97.5	90.3	91.8	75.2
C. Conversion cost:								
Zero cost of conversion	49.0	49.0*	100.0*	1.00*	100.0*	100.0*	100.0*	83.5*
50-percent-of-baseline cost	51.3	46.8*	91.2*	1.09*	99.2*	95.7*	96.4*	80.2*
Baseline scenario	*51.6*	*40.9*	*78.8*	*1.26*	*97.5*	*90.0*	*91.6*	*76.6*
150-percent-of-baseline cost	50.6	23.2*	46.7*	1.99*	92.7*	76.5*	79.9*	71.6*

*Results significantly different from baseline result at 5-percent level.

Note: The baseline experiment used 1,600 realizations of the process; others used approximately 600 realizations.

Table 6
Monte-Carlo Experiments: Variations in Structure of the Model

(Point estimates for results at the end of the process)

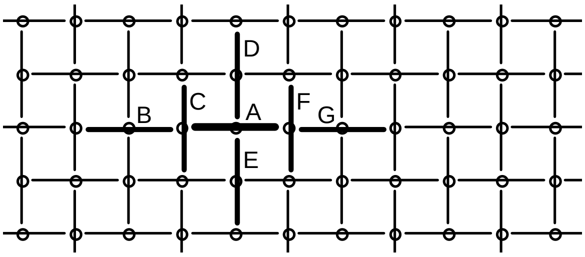
Experimental scenario	Mean proportion of narrow gauge	Proportion of trials resulting in standard	Mean number of net-works	Mean indices of network characteristics Local standard-ization	Continental standard-ization	Network integration benefits	*Ex-ante* effici-ency
	Percent			*Scale of 100*			
Baseline scenario	*51.6*	*78.8*	*1.26*	*97.5*	*90.0*	*91.6*	*76.6*
Random gauge choices	49.9	0.0*	14.0*	49.9*	36.6*	39.4*	([a])
Conversion of random gauges	49.9	30.8*	2.05*	74.2*	65.1*	67.1*	([a])
Single-phase process	47.9	81.5	1.24	97.9	91.3	92.7	78.2*
Four neighboring lines only	48.5	56.5*	1.94*	94.4*	77.1*	80.8*	67.9*
Lattice size of 144 lines	53.9	39.2*	1.83*	92.8*	77.1*	82.1*	79.2*

*Results significantly different from baseline result at 5-percent level.

[a] Statistic not meaningful for this scenario.

Note: The baseline experiment used 1,600 realizations of the process; others used approximately 600 realizations.

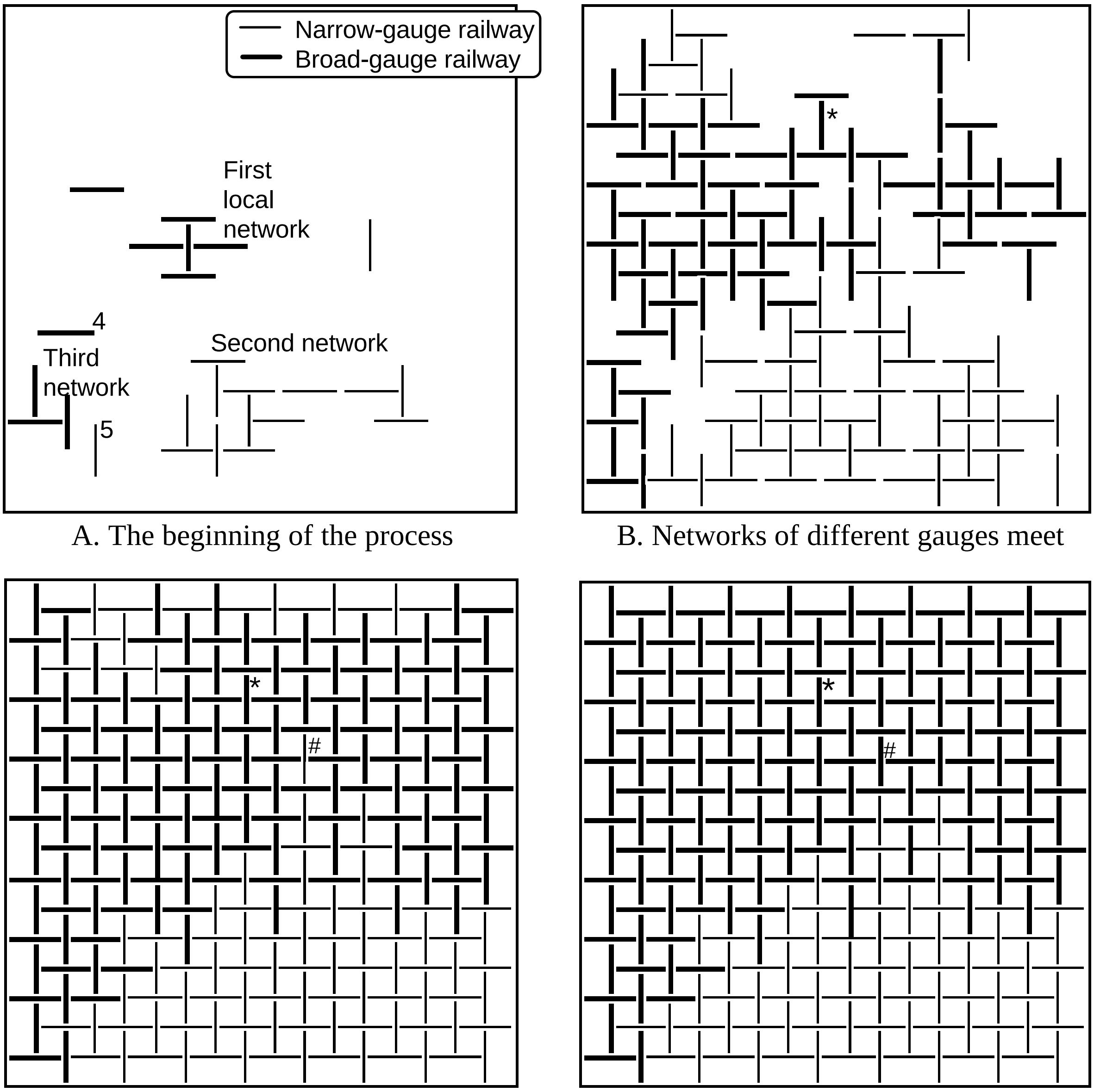
Figure 1
Structure of the modeled network



Note: Circles show the underlying lattice; lines represent local railways, oriented alternately in "north-south" and "west-east" directions. Railway A meets two other railways at each end (B and C, F and G) and two in the middle (D and E).

Figure 2

Numerical Simulation: “Snapshot Maps” of an Evolving Sample Realization



A. The beginning of the process

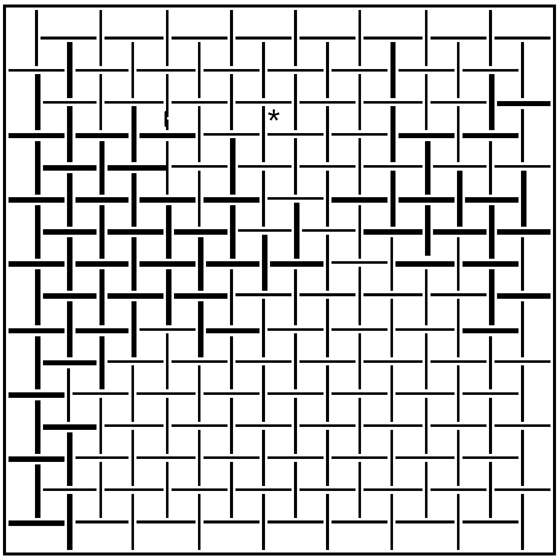
B. Networks of different gauges meet

C. Configuration after construction

D. Final configuration

Note: Maps were drawn by the computer program during the simulation.

Figure 3
Numerical Simulation: Alternate Realization



C. Configuration after construction

Figure 4
Numerical Simulation: Distribution of Results in Monte-Carlo Experiment

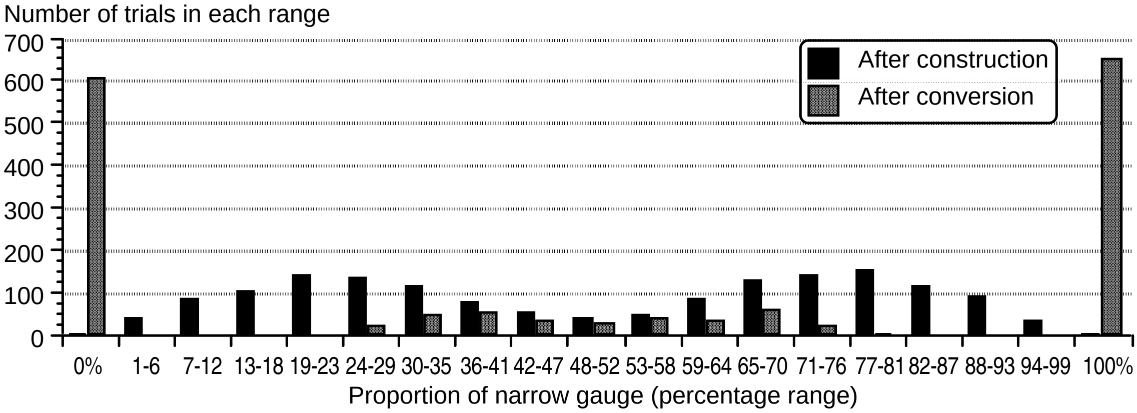
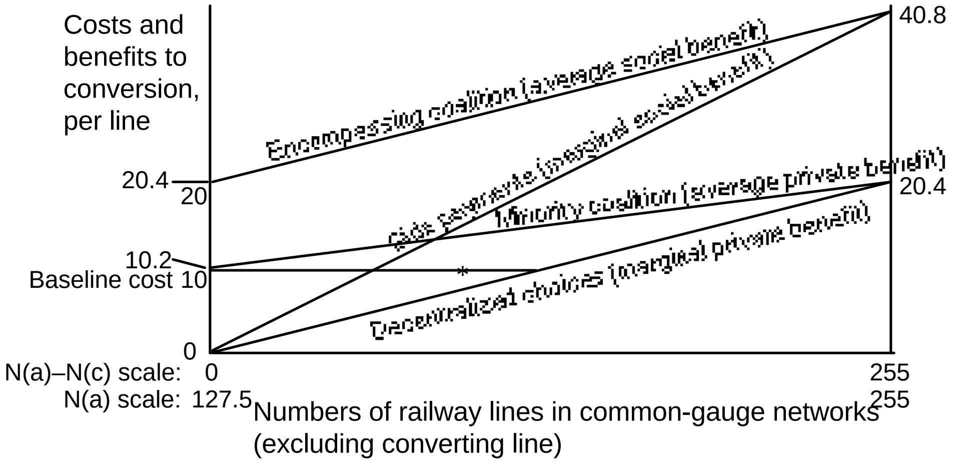


Figure 5
Maximum coverable conversion cost to reduce diversity of gauge, by externality-internalization scheme


Costs and benefits to conversion, per line
Encompassing coalition (average social benefit)
Side payments (marginal social benefit)
Minority coalition (average private benefit)
Baseline cost 10
*
Decentralized choices (marginal private benefit)
N(a)–N(c) scale: 0
N(a) scale: 127.5
Numbers of railway lines in common-gauge networks (excluding converting line)