

Similar Economic Histories, Different Industrial Structures: Transatlantic Contrasts in the Evolution of Professional Sports Leagues

LOUIS P. CAIN AND DAVID D. HADDOCK

Industries that have different structures in Europe than in America can teach useful lessons about industrial evolution. Despite similar initial histories, European professional sports leagues adopted team promotion and relegation, which facilitates much easier entry than is possible in North America, where leagues themselves create new franchises to sell to investors. By contrasting the histories of the English Football League and the National Baseball League, we show that their structures arose from differences in geographic compactness, the entertainment level of games, and territorial monopolies. As the evolution becomes more understandable, the persistence of the intercontinental difference becomes more problematic.

Popular entertainments, including sport, blossomed as the workweek shrank in the mid-nineteenth century. Until 1892 the development of sport was similar on both sides of the Atlantic. Then each continent's pioneer professional sports league, baseball's National League in the United States and soccer's Football League in England, merged with a major competitor. At that moment the two leagues, already pursuing dissimilar paths regarding exclusive team territories, adopted divergent trajectories regarding entry and exit. That is a puzzle. Each league was aware of developments on the other shore, so ignorance cannot account for the radically different modes of operation that emerged. This article explains how, despite similar economic histories, subtle differences in the two countries' economies contributed to substantial differences in costs and benefits—and in league structures.

The Journal of Economic History, Vol. 65, No. 4 (December 2005). © The Economic History Association. All rights reserved. ISSN 0022-0507.

Louis P. Cain is Professor of Economics at Loyola University Chicago, Adjunct Professor of Economics at Northwestern University, and Visiting Professor of Economics at the University of Chicago Graduate School of Business. E-mail: lcain@luc.edu. David D. Haddock is Professor of Law and Economics at Northwestern University, and Senior Fellow at PERC. E-mail: dhaddock@northwestern.edu.

We benefited from the contributions of Ran Abramitzky, Fabio Braggion, John Curran, George Harmon, Michael Hauptert, Charles Hickson, Brooks Kaiser, Lynne Kiesling, Fred McChesney, Robert McGuire, Chiaki Moriguchi, Roger Noll, Kirk Ross, Ian Savage, David Surdam, Stefan Szymanski, Peter Wardley, Samuel Williamson, Todd Zywicki, seminars at DePaul, Loyola, and Northwestern Universities, the 2004 World Cliometrics Conference, the 2004 Southern Economics Association meetings, the 2005 Canadian Network for Economic History, and this JOURNAL'S editor and referees.

Sports economists often contrast the promotion-and-relegation system of European leagues with the closed-entry structure characteristic of the United States and Canada.¹ Few commentators ask why the difference exists, and those few typically fall back on some *ad hoc* cultural explanation. For example, North American team owners are said to be “profit-maximizing businessmen” whereas European team owners are “sportsmen.” Yet even if true, sportsmen must remain solvent. Whether the goal is to maximize profits or to win championships within financial constraints, costs and benefits are important.

The tradeoffs that leagues faced were different in Europe and America, due mainly to their economic geographies. In the short run, intra-league competition is reduced if league members are geographically dispersed, enabling teams to charge higher prices. Dispersal, however, increases travel costs, and, although the Football League operated in a much smaller nation, we argue that it was more strongly influenced by travel costs during its early years. Dispersal also increases the danger that an entrant can find a viable geographic niche between existing teams. Beginning by more thoroughly describing their structures, we elaborate on the puzzle the leagues’ divergent evolution poses before turning to their histories for an explanation.

WHAT IS THE DIFFERENCE IN STRUCTURE?

The peculiarities are not of a sport but of a shore. The National League, founded in 1876, was the first permanent professional team sport league; the Football League, founded 12 years later, was the first in Europe. Each provided paradigms for other leagues on their respective continents.²

North American Leagues

Despite many willing entrants, today’s top-tier North American professional sports leagues maintain a fixed roster of teams—a “closed

¹ Discussion of North American leagues often concerns big-market versus small-market teams, product quality, and so on; for example, Leeds and Allmen, *Economics of Sports*. Insightful literature analyzing various aspects of the promotion-and-relegation system includes Noll, “Economics of Promotion;” Ross and Szymanski, “Law & Economics;” and Szymanski, “Economic Design.” Szymanski and Zimbalist, *National Pastime*, compare various attributes of the alternative systems.

² Important intercontinental distinctions characterize even sports played on both sides of the Atlantic, such as basketball, ice hockey, and soccer. British cricket resembled a North American league until 2000 when it too broke into divisions and adopted promotion and relegation.

corporation,” as National League promoters referred to their structure.³ Only a limited number of teams can compete, and that number is fixed until the incumbents decide to admit others. Each team enjoys a league-enforced territorial monopoly or duopoly.⁴ With three exceptions, teams in the same sport now play in different stadiums.⁵

Starting a new team is expensive in North America. Existing teams, acting through the league, create and sell new franchises for prices vastly in excess of the value of the contracts of players drafted from existing teams or any other tangible property that an entrant receives. A minor league team in the expansion city characteristically moves elsewhere or ceases operation. A rejected entrepreneur’s only recourse is to organize an entire new league. Attempts have been relatively common, but few have succeeded.⁶

Existing teams have relocated with league consent to new cities as far as 2,600 miles (4,100 km) away. An abandoned city has difficulty attracting a replacement.⁷ Although long-distance moves engender acrimony, relocations of several dozen miles are uncontroversial—neither the Angels’ 30-mile (50 km) move from Los Angeles to Anaheim nor the Patriots’ move from Boston halfway to Rhode Island generated much comment.

League rules specify roster size. They forbid loaning players to other teams at the same level, although baseball and ice hockey teams contract with more players than their active rosters can hold, assigning the surplus to “farm” teams that play elsewhere in lower-level leagues. Though teams can trade player contracts, if a cash-strapped owner attempts to capitalize the roster by selling the contracts of a number of team stars, the league will nullify the deals.

³ Rader, *Baseball*, p. 49.

⁴ Only the New York City metropolitan area has had three top-tier teams in a sport at the same time—three Major League Baseball teams from 1903 to 1957 and three National Hockey League teams since 1982.

⁵ Two basketball teams and two soccer teams in Los Angeles share stadiums, as do two New York football teams (though one is eager to move). At least eight miles (13 km) and usually much more separate all other same-sport stadiums. By Stigler’s survivorship standard (*Organization of Industry*, pp. 72–74), the fact that facility sharing, once quite common, has receded to insignificance is instructive.

⁶ Frequent entry combined with a high failure rate implies that entrepreneurs believe incumbent profitability is abnormally high, but no insurgency has succeeded for nearly half a century in any major sport. No team has successfully entered Major League Baseball since 1901 without concurrence purchased from incumbents. Recent attempts by insurgent leagues to enter basketball and ice hockey resulted in a few of those teams eventually being absorbed by the incumbent leagues, but ordinarily all unauthorized entrants are left by the wayside.

⁷ Washington regained a Major League Baseball team in 2005 after being without one for a third of a century; Los Angeles, the nation’s second largest market, has had no professional football team for a decade.

European Leagues

Automatic promotion and relegation is practiced by nearly every European league; strong teams at one level regularly replace the weakest teams from the next higher level. For example, in England the top 20 football teams play in what we refer to throughout the article as the Premier League; three teams from the top division of the Football League annually interchange with three Premier League teams, and so on down.⁸ For illustration, any of the 138 teams presently playing in the five tiers immediately below the Premier League could in principle be Premier League teams in five years or less. Thus the industry is open to entry through superior performance, and any entrepreneur with ambitions to operate a Premier League team enjoys wide latitude to try.

Few (if any) European leagues recognize territorial rights. By North American standards, English professional football teams are tightly packed. Indeed, rival teams at each competitive level sometimes play within walking distance of each other. Measured as the crow flies, the stadiums of three major London teams—Chelsea, Fulham, and Queens Park—are separated by less than three miles (five km). Thirteen football teams playing inside London's ring road have competed in the Premier League at some time, and multiple teams in smaller cities also have reached that level.

A newly formed team simply registers to play in football's lowest tier. It may be geographically near an existing competitor; it needs no one's permission to be there. It then works its way up. Relegated teams do not relocate.⁹ Teams seldom trade player contracts, but they often sell them for cash. The maximum number of players who may play in a game is defined, but the roster from which they can be drawn varies widely across the teams. Surplus players are loaned to other teams in the same or a different tier, or assigned to a reserve team that plays locally. No farm teams play elsewhere.

WHAT IS THE PUZZLE?

When teams agree to form a league, it would seem advantageous for them to opt for the closed cartel now common in North America. On the

⁸ The Premier League was the First Division of the Football League until 1992, when it separated as a legal entity.

⁹ Forced by the deterioration of its London stadium to "ground share" with a nearby rival for more than a decade, Wimbledon's unprecedented 55-mile (89-km) move to suburban Milton Keynes sparked bitter controversy (*Football Grounds*, p. 108). A replacement Wimbledon team formed shortly after the impending move was announced and plays less than nine miles (14 km) from five upper-tier teams, including the defending Premier League champion, http://en.wikipedia.org/wiki/AFC_Wimbledon.

other hand, aspiring entrants, players, and fans are advantaged by promotion and relegation where: teams must compete against nearby rivals to sell tickets and (recently) local broadcast rights; teams must compete against many credible rivals when signing players, given each team's potential upward mobility; incumbents receive no expansion fees from entrants, who may locate literally next door; and teams must bear a larger share of the cost to build and maintain their stadiums.¹⁰ For teams already in a league, the only obvious advantage of the European-style open cartel appears to be enhanced season-end interest in those teams that are attempting to avoid relegation. Why then did the Football League choose an open cartel?

North America

Baseball evolved in the Northeast from informal children's games such as rounders.¹¹ Young men formed social clubs that made arrangements for fields where members played intramurally. In 1862 William Cammeyer made his Brooklyn property available free to local clubs that agreed to let him charge spectators admission.¹² Cammeyer's innovation spread quickly, but soon the teams began to collect the revenues. When receipts surpassed operational expenses, the surplus was used to entice "crack" players away from other teams, though players of more modest talent expected no compensation.¹³ The game spread rapidly among the troops during the Civil War, and among the population at large when the troops returned home.¹⁴ Eventually club all-star "first nines" began to travel to play other clubs' all-stars, but there was no consolidated schedule or championship

A protoleague, the National Association of Professional Base Ball Players, broke away from an earlier, staunchly amateur predecessor in 1871.¹⁵ National Association membership was unrestricted; any team could compete by paying an annual fee; clubs were instructed to ar-

¹⁰ There is less competition among European political units to obtain or retain a given team, because they are not dependent on a league assigning a team to a specific locale. In contrast, North American teams often play virtually rent free in stadiums owned by the local government. If such facilities are not forthcoming, North American leagues readily permit a team to move to a distant city or deny the city a franchise. Okner, "Subsidies"; Baade and Sanderson, "Field of Fantasies"; Knautz, "Franchise Free Agency"; and Poitras and Hadley, "Do New Major League Ballparks."

¹¹ Szymanski and Zimbalist, *National Pastime*, pp. 16–20.

¹² Kirsch, *Creation*, p. 235.

¹³ Within a year of Cammeyer's innovation the *Brooklyn Eagle* reported, "ballplaying has become quite a moneymaking business, many finding it to pay well to play well" (Rader, *Baseball*, p. 27).

¹⁴ Rader, *Baseball*, pp. 17–18; and Eckard, "Team Promotion," p. 124.

¹⁵ Rader, *Baseball*, p. 45.

range, on their own, best-of-five series against each of the others. The best-of-five format meant that a three-game sweep would precede two days without paying fans. Scheduling protocols meant that a western team with an eastern series to play might belatedly withdraw if no intervening opponents with open dates could be found to help defray travel costs. Alternatively, after a team began a trip, an opponent might disband before that part of the itinerary had been reached, leaving the travelers with idle time.¹⁶

The National Association failed miserably to develop an orderly process for teams to contest a championship and balance their books, so the *Chicago Tribune* urged restructuring. The result was the first modern professional sports league, “a new baseball association based on territorial monopolies and a restricted number of strong, financially healthy clubs.”¹⁷ The National League excluded weak teams that had imposed external costs on their fellow National Association members. To the modern observer, the league’s centrally determined schedule may appear an obvious way to avoid many of the National Association’s failings, but it was a major innovation at the time, one that rivals would quickly imitate. In order to facilitate extended, multigame trips that economized on travel costs, the new league selected its member clubs with an eye to geography—four in the old Northwest (Chicago, Cincinnati, Louisville, St. Louis) and four in the Northeast (Boston, Hartford, New York, Philadelphia). The major leagues then remained within the northeast quadrant of the United States for the next 82 years. Only one club could represent a city. Until 1892 the territorial monopoly that came with league membership represented a transitory right, and nonleague teams often supplanted solvent clubs between seasons.¹⁸

Of the several early attempts to circumvent National League exclusion, only the American Association survived for more than a season. Because three founding clubs had previously played in the National League, they understood the advantages of a coordinated schedule and a one-team-per-city limit. During 1882, the American Association’s first season, none of its teams played in a National League city. That segregation ceased the next season when each league invaded one rival market. Over the next ten years, the two leagues competed head-to-head in eight cities. The National League characteristically won those confron-

¹⁶ Twenty-four different teams competed in the National Association, though never more than 13 during a season. There was at least one midseason failure every year. Eckard, “Team Promotion,” p. 126; Nichols, *Major League Franchise Information*. During the second season, two teams played 54 games while one played only nine. Only three clubs fielded a team during each of the five seasons.

¹⁷ Goldstein, *Playing for Keeps*, p. 147.

¹⁸ Eckard, “Team Promotion,” p. 131.

tations, occasionally pirating an American Association club in the process.¹⁹ The coup-de-grâce came after the 1891 season when the eight National League teams invited four from the American Association to amalgamate with them to form a 12-team league.²⁰

The American Association's demise marked the end of city entry into the National League until the league began to transfer franchises 62 years later, and it marked the end of team entry into that league until its first expansion franchises were sold 71 years later.²¹ A new National League constitution converted transitory membership into a near permanent right by forbidding involuntary removal of teams that met specified conditions. By 1900 the league had bought out four member clubs (three of them had become virtual farm teams of other league members) and was back to eight, marking a permanent end of team (as opposed to city) exit from today's major leagues.²² Putting aside its territorial monopolies, National League club turnover before 1892 resembled the system the Football League employed until that same year. It was the closest approach North American professional sports leagues would ever make to the promotion-and-relegation system that is now commonplace throughout the rest of the world.

The only long-lasting challenge to the National League came from the minor Western League. This upstart developed quality teams in small markets, and then in 1901 moved them into three of the four cities that the National League had just abandoned, while going head-to-head in three others. Renaming itself the American League, the newcomer refused to recognize the National League's system of reserving the services of a player to the team with whom he first contracted.²³ Over the next two years the American League transferred two more franchises and was in direct competition with six National League clubs in the nation's five largest markets.²⁴ Perceiving a credible threat, the National

¹⁹ The teams now known as the Dodgers and (ironically) the Pirates were pirated from the Association. Comprehensive major league statistics, including those of defunct "major" leagues, are available at <http://baseball-almanac.com/>.

²⁰ Eckard, "Team Promotion," p. 130. As Eckard points out, the National League-American Association "merger not only eliminated competition between leagues, it apparently also eliminated competition among clubs for league membership" (p. 131). He argues that the closed cartel was adopted in baseball "by owners seeking to protect themselves from that particular type of competition" (p. 146) by attempting to confine nonleague teams to permanent minor league status. Our article explains why the Football League admitted competitors rather than trying to force them into a permanent subordinate status.

²¹ Eckard, "Team Promotion," p. 122.

²² Eckard, "Team Promotion," p. 143.

²³ The reserve system initially covered only four players per team, then gradually expanded, but it did not cover the entire roster until later. Such a reserved right remained in force even after the team's contract with a player expired, was alienable to other teams, but monopsonized the player's baseball playing ability.

²⁴ In 1898 the city of Brooklyn became a borough of New York City, creating a 60-year-long technical exception to the National League's one-team-per-city rule.

League grudgingly accepted the junior circuit as co-equal via the Major League Agreement of 1903. The U.S. Supreme Court would soon legitimize baseball's closed cartel by granting an antitrust exemption.²⁵ The major leagues then stood pat for half a century, retaining precisely the same teams and remaining in exactly the same locations.

Europe

The evolution in England was similar. Informal games resembling football trace back millennia, but interscholastic competition between English schools sparked the first rule codification, the Cambridge Rules of 1848.²⁶ Graduating students formed clubs in their home cities. Eventually all-star elevens began to compete. As in early baseball, there was no central schedule, no championship.

In 1863 the Football Association was formed to refine the code and resolve disputes, incidentally severing soccer from rugby. Eight years later—the year that baseball's National Association was formed—the Football Association established the FA Cup, an open-entry knockout tournament that still exists.²⁷ The following year, the Football Association initiated international matches among England, Ireland, Scotland, and Wales.²⁸ Well before the Football League emerged, the FA Cup and the internationals had grown popular, influencing the continents' divergent evolution—whenever the Football League has tried to distance itself from nonleague members, the Football Association has threatened to exclude league teams from the FA Cup competition and their players from the internationals. Were they to do so, the disciplined teams would find it difficult to hire players.

During the last quarter of the century, working-class men increasingly played football and imitated the club structure. Oxbridge graduates were over-represented both in the Football Association administration and the football-playing population in and around London. Many refused to play against anyone who was paid to play. Industrial worker-

²⁵ *American League Baseball Club v. Chase*, 149 N.Y.Supp. 6 (1914); *Federal Baseball Club of Baltimore, Inc. v. National League of Professional Baseball Clubs, et al.*, 259 U.S. 200 (1922), affirming 269 Fed. 681 (App. D.C. 1921), which had reversed the trial court. The *Federal Baseball* decision “appears strikingly out of touch with its subject matter . . . [T]he decision, viewed as an effort to characterize the actual legal and economic character of Major League Baseball in 1922, was remarkably myopic, almost willfully ignorant of the nature of the enterprise” (White, *Creating the National Pastime*, p. 70).

²⁶ Young, *British Football*, pp. 74–76; and Harris, *Sport in Britain*, p. 107.

²⁷ The Football Association's cup tournament was copied throughout the soccer world, including the United States, always separate from, though simultaneous with, each nation's league season.

²⁸ Even today, the United Kingdom's principal components compete independently in international competitions such as the World Cup.

players from the Midlands and the north of England, however, could ill afford trips to distant games.²⁹ In 1885, three years before the Football League formed, the Football Association reluctantly voted to allow payment for “broken time,” or wages lost while playing and traveling.³⁰ Clearly some players had been paid surreptitiously even before 1885, because a number of working-class football clubs had been suspended or otherwise punished by the FA for that infraction.³¹ Because, after 1885, some locales generated more revenue to pay broken-time compensation to semiprofessional players and even more to stars, the variance of skill levels across teams increased, and with it the number of unappealing one-sided contests. Teams began to seek parity, in other words to seek opponents of a similar caliber.

The Football League formed in 1888 when Birmingham’s Aston Villa club proposed that a few top teams annually play one another home and away, and 11 others agreed.

“[G]ood gates had become imperative” to the leading clubs, but the multiplicity of local cup competitions, often with one-sided early rounds, could not guarantee this and they also interfered with more remunerative friendly fixtures against other major teams.³²

The Football League’s founders were familiar with the National League and imitated its coordinated schedule, but, after debating the matter, rejected territorial restrictions. Six Lancashire teams, three in-and-around Birmingham, and three in-between comprised the initial league.³³

The FA’s rule change means that Football League clubs were most likely comprised of a mix of semiprofessionals and full professionals, though we have been unable to find evidence regarding the speed with which professionalism spread.³⁴ It is plausible that, on the best-

²⁹ London’s well-heeled amateur Old Etonians traveled to Lancashire to take on working-class Darwen. “The result was a draw, but Darwen were unable to travel to London for the replay. The Etonians offered to pay their fares, but the Darwen players simply could not afford the loss of wages . . .” Harris, *Sport in Britain*, pp. 111–12.

³⁰ Harris, *Sport in Britain*, p. 112.

³¹ Vamplew, *Pay Up*, pp. 191–92.

³² Vamplew, *Pay Up*, p. 125, quoting Aston Villa’s William McGregor.

³³ Aston Villa (Birmingham), Blackburn, Bolton, Everton (Liverpool), and West Bromwich (Birmingham) were in the Premier League during the 2005/06 season, with Burnley, Derby, Notts County (Nottingham), Preston, Stoke, and Wolverhampton in lower divisions. The twelfth team, Accrington, resigned in 1893, then folded two seasons later, <http://accringtonstanley.has.it> (Derby played until 1997 in the Baseball Ground, named for one of the facility’s earlier uses.)

³⁴ We use “professional” to mean a player who is not merely paid, but paid enough to support himself during the season without an additional job. A semiprofessional, while paid something for playing, holds a regular job during the season and plays during off-hours. A person paid to play for an industrial team used for promotional purposes by a nonsports firm is a professional, unless the player is expected to participate meaningfully in that firm’s normal business activities during the season.

supported teams, the most skilled players earned enough to give up outside employment. Even today a gradual transition from semi-professional to full professional teams in English football continues. At present about half of the 22 Conference teams—those in the fifth tier whose strongest members are annually promoted into the Football League—employ semiprofessional players. At its inception, the Football League was not a national league in either name or fact; it kept games near its semi-professional players' main jobs.

As had happened in America after the National League provided its seminal example, the advantages of the Football League's coordinated schedule quickly became apparent to other English clubs. The Football Alliance, a rival league with a similar geography, sprang into existence the following year.³⁵ From the beginning, the Football League required annual reapplication by the four lowest-ranking clubs.³⁶ In its third season the league replaced Stoke with Sunderland, an even more northerly team that it pirated from the Football Alliance. Two years later the league initiated its now characteristic divisional structure upon absorbing the rest of the Football Alliance.³⁷ Annually interchanging a few teams between divisions maintained parity more systematically, reapplication being required only of the lower division's bottom teams. Thus in 1892, the precise year that the National League abandoned what Eckard termed its "quasi team-promotion system," firmly and finally closing its cartel, the Football League began crystallizing its once similar structure into the comprehensively open system one finds today.

In 1893 Arsenal became the Football League's first London team, though it languished in the second division until 1904. The league only moved south with alacrity in 1920 when the upper division of still another rival, the Southern League, which had formed in 1894, was absorbed as a third division.³⁸ The following year the league added 20 small northern teams. From 12 teams to 86, the Football League's average expansion rate between 1888 and 1921 was well over two teams per season. Over a comparable period following National League forma-

³⁵ Robinson, *Non-League Football Tables*, p. 4.

³⁶ The stated motivation was to assure that league teams were the best available, but rail connections also weighed heavily. "What has been the first point urged in any application for membership . . . ? The ease with which visiting clubs could get to the new centre and away from it." Journalist James Cotton, as quoted in Taylor and Coyle, "Election," p. 9.

³⁷ Arnold, "Rich Man, Poor Man," p. 52. By the turn of the century, Alliance members Birmingham, Grimsby, Manchester City, Manchester United, Nottingham Forest, Sheffield United, and Sheffield Wednesday all had reached the top division for at least one season.

³⁸ Arnold, "Rich Man, Poor Man," p. 52; Robinson, *Non-League Football Tables*, p. 9. Season-by-season Football League records are found at http://www.the-english-football-archive.com/football_league.htm.

tion, the major leagues expanded from eight teams to 16, less than a quarter of a team per season.

The Inadequacy of a Cultural Explanation

The difference in league structures is sometimes said to reflect a difference between European “sportsmen” and American “profit-maximizers.” Certainly the founders of the Football League did not manifest as much commercial zeal as the National League’s founders.³⁹ That is explainable, however, by the different economic environments the teams faced. Innovators do not dictate to their economy; an economy selects viable innovations from among the possibilities that are offered by its array of innovators. In the daunting environment in which baseball operated, teams faced substantially higher costs, which forced them to find ways to generate commensurate revenues. The environment in which football operated permitted a goal of winning championships to be primary, but football could not have survived as its costs increased without recognizing its financial constraint. In either environment, financial considerations were (and remain) crucial.

Empirically, history reveals that the first teams on both shores originated from informal amateur recreations, sportsmen’s teams run for and by the players. Prior to league formation on both continents, when spectators betrayed a willingness to pay, a few teams began to use some receipts to entice superior players to join. The strongest teams absorbed more player time during the season and paid them accordingly. For a variety of reasons, players’ salaries and travel costs were higher in the United States, so teams were forced to focus on finance at an earlier date than was necessary in England. This focus meant professional management grew in importance more rapidly in the United States.⁴⁰ Methodologically, distinguishing a financial goal from a financial constraint would seem to rely on an unobservable phenomenon—entrepreneurial motivation—which would render the distinction untestable and, as will be shown, unnecessary.

³⁹ For example, Szymanski and Zimbalist point to the different personalities of William Hulbert, the National League’s catalyst, and William McGregor, who played a similar role for the Football League. See Szymanski and Zimbalist, *National Pastime*, pp. 44–47. There is a sports-related example of a more general claim that there were innate attitudinal differences between British and American businessmen. See Chandler, *Scale and Scope*.

⁴⁰ On the transition, see Goldstein, *Playing for Keeps*, regarding baseball; and Vamplew, *Pay Up*, regarding soccer.

WHAT RESOLVES THE PUZZLE?

The issues to be resolved, then, concern divergent league policies toward market division and toward entry. Why, despite similar origins, did the Football League's founders reject the territorial monopolies that appeared to benefit the National League? Why did the Football League open its haphazard election system into formal promotion and relegation in the same year that the National League was barring the door against its own previous haphazard system of club entry? Our explanation focuses on geographic compactness, its impact on wages and travel (both for teams and their fans), on the cost side, and on the excitement level of games and ability to maintain territorial monopolies on the benefit side.

Compactness

Uncoordinated scheduling proved easier in densely populated England than in sparsely populated America, because teams in the United Kingdom had many more nearby opponents among which to choose. Even top-flight English clubs could ordinarily arrange games against comparably strong opponents at locations that permitted round-trip travel within a day, requiring little disruption of the work schedules of the fans as well as those players who continued to hold outside employment.

A similar environment existed for the lesser clubs in urban America, but the best teams discovered that they could either travel short distances to play generally one-sided games against inferior opponents before few fans, or they could travel long distances to meet comparable teams before large crowds composed almost entirely of home-team fans. Long-distance travel disrupted players' work at their full-time jobs, so U.S. baseball's top teams became fully professional much earlier than any English football club. Once a team had to travel so far that its players could not hold outside jobs during the season, the marginal cost of an increase in travel distance diminished. Though baseball's *total* travel costs were much higher, their *marginal* travel costs with respect to distance were lower; thus, they became less sensitive to distance than were U.K. football teams.

Old England occupies about three-quarters as much area as New England, as shown in Figure 1, where the countries are drawn to the same scale. According to the 1890 U.S. and 1891 U.K. censuses, however, Old England contained roughly half as many people as the entire United States. Even today England and Wales hold more than a fifth of the U.S.

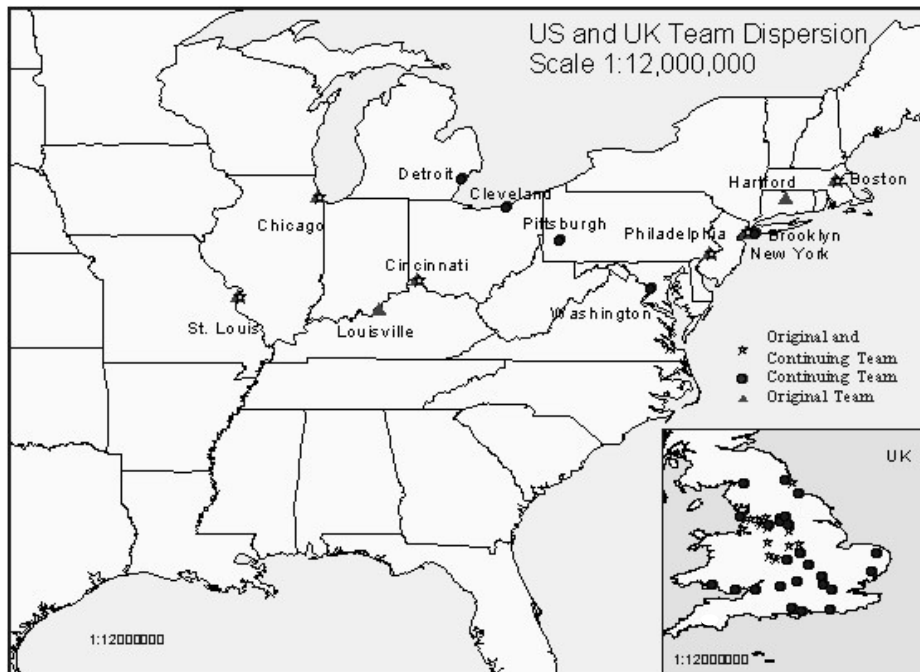


FIGURE 1
U.S. AND U.K. TEAM DISPERSION c. 1950

Source: Authors' data with the assistance of Professor Brooks Kaiser.

population in an area the size of Michigan. The Football League covers only a middling area among major European soccer leagues, barely half that of Italy, but even the league in France, Europe's most extensive, covers less than half the area of the northeast quadrant of the United States, where the National League began play.⁴¹ Due to their greater population densities, European nations were blessed with a denser rail network than the United States.

To illustrate that the requisite travel would have rendered supplemental in-season jobs less viable for American than English players, we contrasted the geography of the National League with that of the Football League upon their respective inaugurations. The average separation between the eight original National League teams was ten times that between the 12 original Football League teams—595 miles (958 km) versus 59 miles (95 km).⁴²

⁴¹ France has an area between those of California and Texas, and its first professional football championship was not contested until the 1932/33 season. Although Russia is the largest country in Europe, the area spanned by its football league is limited. In fact, nearly all noteworthy Russian soccer teams play in Moscow.

⁴² An almost identical ratio applies to the separation in today's leagues, though with air travel available and fully professional top-echelon players in both countries distance is less important.

Over the history of the Football League, the furthest two teams have ever traveled for a league game is 410 miles (660 km), the distance between Plymouth and Newcastle. That trip is about the same as that from Boston to Baltimore, or Los Angeles to San Francisco. It is less than the *average* separation between the original National League teams, and less than a third of the average separation between today's major league teams. Manchester and Liverpool are 35 miles (55 km) apart. Either one is less than 95 miles (155 km) from Birmingham, roughly the distance between Chicago and Milwaukee. These three English cities are closer than the minimum separation between National League teams in 1876, the 101 miles (163 km) between Cincinnati and Louisville. English compactness led to lower travel times and thus lower total (as distinct from marginal) travel cost.

Travel time is a reasonable proxy for the average player's opportunity cost, and the time that would have been required during the late nineteenth century to travel round-trip to another city and play an average game can be estimated. Assuming that compensation foregone per hour was equivalent in baseball and football, we took the intercity mileage, assumed average train speeds of 35 mph (55 km/h), then added six hours (three hours for getting to and from the stations at each end, dressing, warm-ups, undressing, and treating injuries, plus a generous three hours for playing the game).⁴³

None of the 66 city pairs in the original Football League would have required as much as 14 hours for that round trip, and 54 (82 percent) would have required less than 12 hours. In contrast, among the 28 original National League city pairs, 24 (86 percent) would have required more than 14 hours, and none could have been completed in less than 12 hours. Indeed, nearly two-thirds of the National League round trips would have required more than 24 hours, nearly half would have required more than 48 hours, and a round trip between St. Louis and either Boston or Hartford would have taken almost three full days.

The average distance between today's 26 major league cities is 1,317 miles (2,120 km) whereas the average distance between the 45 cities shown in Figure 1 with teams that have played Premier League football is 131 miles (210 km). For present purposes we drew no distinction between continuously operating teams and reorganizations or replacements for failed clubs that had operated at the same site. Four former Premier League teams or their successors began the 2005/06 season below the Football League level.

⁴³ The assumptions are based on samples from the timetables in *Travellers' Official Railway Guide* for the United States (1899) and *Bradshaw's Guide* for Great Britain (1882). Foregone earnings swamp direct travel costs, but account book information assembled by Hauptert and Winter, "Pay Ball," suggests that in 1915—the first year for which data are available—the New York Yankees' direct travel costs were equivalent to the annual salaries of five players.

The average round trip in the Football League (including game time) was less than half a day and well under the National League's minimum. Plymouth didn't play a league match against Newcastle until 1934, but even in 1888 that round trip would have required only 30 hours, well under the 40-hour average in the original National League.

Because strong competitors were nearby, English football teams could leave for most away games Saturday at noon (the end of the British workweek until the 1950s), play a game against a closely matched opponent during the afternoon, and return home Saturday evening. Separating the teams by enough distance to form monopolies would have meant paying the players more, because the semiprofessionals among them would have experienced more broken time. Separating the teams into territorial monopolies would have prevented many fans from traveling with a team, reducing revenues. European fans have always attended away games in large numbers, and from the beginning most Football League games were scheduled to accommodate them. In contrast to all North American sports, night games remain rare in English football because so many of them would require visiting fans to spend a night away from home. Thus, when the Football League formed, there was good reason to eschew territorial monopoly.

By contrast, a quality baseball team had to travel so far that its entire roster became fully professional of necessity; no player could maintain outside employment during the season. Few supporters could travel to away games, so attendance was almost exclusively local fans. When the National League formed, there were few potential cost savings to be netted against territorial monopoly advantages. With fully professional players on both continents and the availability of air travel, the travel cost difference has fallen substantially, but even today in the United States, visiting teams and fans generally must incur at least an overnight stay.

Excitement Level of Games

The excitement level of games is a function of parity and variety/rivalry. The literature has discussed the benefits of parity extensively—any league's success depends on staging games with uncertain outcomes.⁴⁴ If outcomes become too predictable, games become boring, and fewer fans attend. We characterize parity as having positive, but decreasing, marginal value. Under normal circumstances, sports fans prefer a greater than 50 percent probability that their favored team wins, but at-

⁴⁴ See, for example, Canes, "Social Benefits"; and Szymanski and Zimbalist, *National Pastime*, pp. 170–94.

tendance wanes when the likelihood of a win becomes too great.⁴⁵ Consider one team in a two-team market. The other local team offers low travel costs, but two teams playing repeatedly expose parity differences. So, variety is also a virtue, again with positive, but decreasing, marginal value. Variety and rivalry are negatively correlated. If rivalry also has positive, but decreasing, marginal value, an optimal league size exists where adding a team leads to a loss of passion that overwhelms the weak benefit of added variety, but the increased rivalry that results from subtracting a team fails to compensate for the ennui from repetition.

To produce well-played games with uncertain outcomes in the sparsely populated, relatively rail-poor United States, the National League restricted membership to one outstanding team in each of eight cities that spanned 1,207 miles (1,942 km). For the sake of parity, the league might have begun with even fewer teams, but that would have provided too little variety.⁴⁶ Although the Football League began in England's midsection with 12 teams spanning 126 miles (203 km), parity tended to be higher than the National League's had been at a similar age (see Figure 2).⁴⁷

Because comparable, high-quality opponents were much denser, variety was less costly to English teams, so, until reaching long-run equilibrium, any Football League division would be predicted to have more teams than would baseball. That requires a relatively strong assumption for cross-section analysis—omitted variables have similar impacts on baseball and football—though that assumption is unnecessary for time-series analysis. Testing that hypothesis is trivial. The Football League began with a larger membership and grew to 28 teams within four years. Despite the much larger U.S. population, well over a century passed before 28 teams were eligible to compete for the World Series championship. Even today there are only 30 such clubs, and only 22 have won a title. One hundred thirty-one teams have played in the Football League, 20 are in the Premier League and another 72 in the three

⁴⁵ Surdam, "American 'Not-So-Socialist' League," found that revenue sharing benefited the New York Yankees in their post-World War II glory days because they were a better road than home draw. While New York attendance waned, fans outside of New York City took advantage of their limited opportunities to see the Yankees' star players, despite the modest probability that their home team would win.

⁴⁶ Cincinnati won less than one game of every seven they played during the inaugural season, whereas Chicago won nearly four of every five. Chicago swept the ten games it played against Cincinnati by an average score of 11.4 to 3. By season's end the games must have engendered little anticipation in either city.

⁴⁷ The idealized standard deviation, devised to study baseball (Quirk and Fort, *Pay Dirt*), is a commonly used parity measure that adjusts for different season lengths. In "Measuring Parity," we explore the subtle errors that arise when applying that alternative to soccer and other sports that permit games to end in ties.

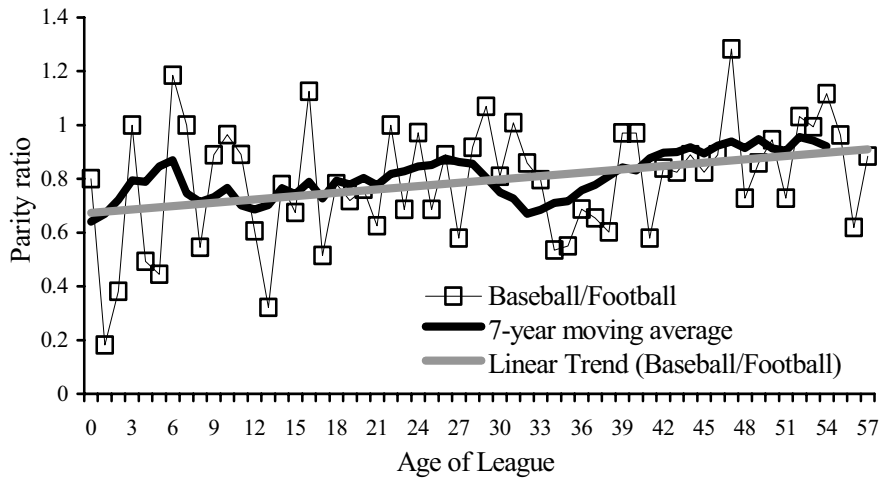


FIGURE 2
PARITY OF BASEBALL AND FOOTBALL, 1888–1956

Source: Authors' calculations from official final standings for each year.

Football League divisions today. Sixty-three have reached the Premier League at some point—an average of more than 29 separate teams during any five-year span since the Football League added a fourth division in 1921—and 23 have won its championship. From the birth of the Football League in 1888 through the leagues' seasons that began in 2005, the inverse Herfindahl yields an equivalent of 35.9 Premier League teams, but only 23.5 Major League Baseball teams.⁴⁸

If one grants our strong assumption that omitted variables have similar impacts, parity was less costly in England at the outset, so the Football League would have begun as the more balanced league. Even without such an assumption, falling travel costs would raise parity on both continents, though more strongly for baseball where the initial

⁴⁸ In distinction to the Herfindahl-Hirschman index, which sums squared *percentage market shares* across all of an industry's firms [$(25)^2 = 625$, for a firm commanding a quarter of its industry], the Herfindahl sums squared *fractional market shares* [$(1/4)^2 = 1/16$, or more conveniently, $(0.25)^2 = 0.0625$]. Though the alternative measures of industrial concentration differ by a constant factor of 10,000, Stigler (*Organization of Industry*, p. 31) showed that the Herfindahl's inverse is equivalent to the number of firms of identical market share that would have produced that same Herfindahl measure, offering an intuitive view of what concentration differences imply. For our computations, a given team's market share is the number of years that it has competed at the top level divided by total team-years of competition for all teams that have competed at that level since the Football League was established. If the computations include the Players League, the Federal League, and the American Association, baseball's inverse Herfindahl increases by one team. Either way, the ratio of football to baseball rounds to 1.5 to 1; American baseball has been substantially more concentrated than English football.

costs were greater. Testing the parity prediction is more challenging than testing the variety prediction. The parity measure we employ is the proportion of clubs that win between 40 percent and 60 percent of the maximum points (football) or games (baseball) available to a team in a season.⁴⁹

Figure 2 contrasts parity in the Premier League and Major League Baseball at similar league ages (the effect remains when we use chronological dates that begin each series in 1888). We end the analysis with the 1955/56 Football League season due to several exogenous events that seem to have affected football parity adversely—court decisions that eliminated maximum salaries, broadcasting’s advent as an important revenue source, and the inauguration of continent-wide competitions. It is unclear *a priori* how salary cap elimination or broadcasting should be expected to alter football parity vis-à-vis baseball, as comparable phenomena occurred in North America. The continental competitions, however, would predictably reduce it.

In the continental competitions—principally the Champions Cup (subsequently the Champions League) and the UEFA Cup—Europe’s best few teams according to the prior season’s results play against each other. That provides a bonus for excelling—a lucky few teams play additional games against similarly skillful foreign clubs. Of necessity being interwoven between weekend league games, the continental contests are played on mid-week evenings against unusually distant clubs. Because fewer visiting fans will attend, those extra games are more valuable to teams with strong local followings, and, as of the 1956/57 season, they had an added incentive to invest in team quality. Perceiving little chance to “play in Europe,” the many teams that rarely play in football’s upper divisions, and then for only a few seasons at a time, were given little added incentive to improve.⁵⁰ The number of teams admitted and the number of games played by the median team—and thus the influence of the continental competitions—have increased over

⁴⁹ For consistency we adjusted the data after 1981 when wins in football became worth three rather than two points. Draws have always been worth one point, but are no longer permissible in American baseball. Because we do not adjust for different season lengths, there is a downward bias in our measure of parity for football relative to baseball.

⁵⁰ Noll, “Economics of Promotion,” p. 199, discusses the incentives that marginal clubs have to pursue such a strategy, becoming what may be called “yo-yo” teams. We ran preliminary regressions of parity against league age that included dummy variables for both the Champions Cup/Champions League and the UEFA Cup. At conventional confidence levels, those regressions suggest that both had a negative impact on parity in the top Football League division, whereas the Champions Cup/Champions League had an even stronger negative impact on the Premier League, though the UEFA Cup had little. Chow tests lead to strong acceptance of the hypothesis that there was no break between 1955 and 1956 for Major League Baseball, weakly rejects the hypothesis for the Premier League, and strongly rejects it for the top Football League division.

the years.⁵¹ A regression of the ratio of Major League Baseball parity to Premier League parity against league age suggests convergence, the trend line's ratio increasing from 0.673 at the start to 0.910 at the end. The statistical confidence that the increase of 0.004 per year differs from zero approaches 99 percent ($p = 0.015$).⁵²

The National League was pinched on the variety margin despite including a few fringe members, as it was pinched on the parity margin because it included them. To produce a more attractive product that balanced variety with parity, the National League rejected open entry. Assuming that the league started with the appropriate teams, potential applicants either were too far away (say, in San Francisco) implying travel that would be too time-consuming and costly, or they were in a market that was thought to be unable to support a suitable team.

The National League was prepared to expel teams that did not measure up, but tried to discourage autonomous exit by those that did. During the early league history, a league club might play as many "exhibitions" against nonleague foes as it played league games, but playing such games in league cities was banned from the outset. As the number of league games that comprised a season grew, exhibition games were virtually eradicated. Consequently, any high-quality team that left the league by choice or expulsion experienced difficulty scheduling worthy opponents, and often failed.⁵³ In contrast, frequent nonleague games remain staples of English football, institutionalized in the FA Cup, continental competitions, and various others.⁵⁴ The Football League could be more blasé about turnover because, with greater team density than the National League, a defecting team was relatively easy to replace.

The Football League could also be more relaxed about the danger that an owner would weaken his team by capitalizing the contracts of star players. The problem would be dealt with promptly as the weakened team would be relegated to a division more in keeping with its reduced ability. If a North American team attempts to capitalize its roster,

⁵¹ Liverpool earned about as much from broadcast rights for its 2005 Champions League games and performance bonuses as from broadcast rights for the team's Premier League season; Chelsea and Arsenal did almost as well. Though eliminated early, Manchester United nonetheless earned nearly half as much as Liverpool. Gate receipts for the four teams' home Champions League games were additional. British Broadcasting Company, *Champions League Winners*, p. 2. UEFA Cup earnings would have been more modest. The majority of Premier League teams participated in neither competition and thus earned nothing.

⁵² Statistical analysis based on the idealized standard deviation leads to identical conclusions to those reported, providing it is appropriately corrected for the many tied games in football.

⁵³ Eckard, "Team Promotion, p. 143.

⁵⁴ A participant in the Champions League might play 19 additional games during the 2005/06 season; its Premier League season consists of 38 games.

however, the impact on parity, and thus the external cost on other teams, could prove long-lived.

Territorial Monopolies

In sharp contrast to the nascent Football League, National League founders seized the opportunity to create territorial monopolies from the outset. Initially the league replaced clubs between seasons when that suited the interests of those that remained, but no team has ever been permitted to locate in a city that is already the home of another National League team.⁵⁵ So from its birth the National League was closed to a degree that the Football League never has been. Following the American Association's demise, entry into the National League—even into a nonleague city—became extraordinarily difficult.

Monopoly successfully achieved, even if for meritorious purposes, may provide market power. When the first additional market grows to a size adequate for a major league caliber team, entrepreneurs face an expensive choice—organize an entire league, as the American League did, with members who either must face entrenched franchises or be sited in initially substandard markets (risking piecemeal invasion as a result of major league expansion or franchise movement as those markets grow); or offer to purchase a franchise from the league.

A European team of any level, coordinating with no other club, need only keep winning to reach the top. The Football League's mergers with the Football Alliance and Southern League as well as the more recent emergence of the Conference illustrate that credible alternative leagues are easy to form in England's compact, densely populated environment. When the better teams became fully professional, the multidivisional promotion-and-relegation system was firmly entrenched. By 1921 the Football League already had 86 teams, and 33 had played in the Premier League. It seems that the Football League could not have maintained an effective closed cartel, and it never tried.

Assume that the Premier League had decided to adopt North American-style territorial restrictions for the 2005/06 season. Further, assume each metropolitan area was limited to a maximum of two teams within a

⁵⁵ Only two of the original eight clubs—those known today as the Cubs and Braves—remained in the league as its sixth season began. Two strong National Association teams, the Whites and the Athletics, had played their home games in Philadelphia. Incorporating both would have improved parity, but defeated a territorial monopoly in Philadelphia. Only the Athletics were admitted to the National League. The American League has followed the same policy regarding team territories, though neither league has consistently respected territories of the other league, resulting in a few duopoly markets in the largest metropolitan areas.

30-mile (48-km) radius.⁵⁶ This means that only two of the six London teams in the league could stay there; four would have been forced to move to much smaller cities if they wanted to remain in the Premier League. Indeed, considering the number of teams separated by less than 30 miles in other parts of the country, nine of the 20 Premier League teams would have had to move to meet the territorial restrictions. Rather than move, there would be an incentive for the potentially displaced teams to remain in place and play in the Football League. Thus, the likely result would be two leagues of similar ability and credibility with no alteration in any team's capacity to control a territory. As the National League discovered when the American League moved into its territory, when faced with credible insurgents, cooperation makes more sense than exclusion.

But why did the initial teams institute the promotion-and-relegation system that permitted so many competitors to gain a foothold in the first place? To understand that, it is necessary to recognize that the interests of league teams are heterogeneous. Expansion of the number of teams in a single league became unattractive once the marginal values of variety and rivalry were offsetting, so a divisional structure was adopted. Random year-to-year fluctuation in team quality meant that promotion and relegation within a divisional structure would maintain a higher level of parity. Teams with a strong expectation of playing in the Premier League would benefit, though weaker league members might not. English football is definitely fluid, but a core has enjoyed frequent Premier League membership, and, while doing so, promotion and relegation has provided them with greater parity.⁵⁷

As long as travel costs are significant, our analysis predicts football team survival in smaller markets than are necessary to support baseball teams and, collaterally, English metropolitan areas that possess more teams for a given population. Given England's size, virtually every game is against a neighbor (at least from a North American perspective). The lack of territorial monopoly increases price elasticities in European football, but games draw fans from two markets so the more elastic demand curve is shifted to the right.⁵⁸ In con-

⁵⁶ A 30-mile radius seems conservative. The Orioles asked for compensation for dilution of territorial rights due to Major League Baseball's relocation of the Montreal baseball team to Washington, 40 miles (64 km) from Baltimore. San Jose's mayor wants a major league team for his city, but the Giants were given territorial rights to San Jose upon moving to San Francisco, 50 miles (80 km) to the north.

⁵⁷ No team has played its entire history in the Premier League, but nine teams have played there for more than two-thirds of its 107 seasons (there was no competition during either world war).

⁵⁸ Though European football leagues enforce no territorial monopolies for their teams, FIFA, the world governing body, enforces territorial monopolies for leagues. A league in one nation

trast, most baseball attendance is drawn from the territory over which a team has a monopoly.

Intercountry differences in market size are difficult to quantify because the United Kingdom and the United States do not define metropolitan areas in analogous ways; the “conurbations” as defined in the 2001 U.K. Census cover substantially smaller areas than the “combined statistical areas” (CSA) defined in the 2000 U.S. Census. London’s conurbation appears to be about as populous as the San Jose-San Francisco-Oakland CSA, the fifth largest in the United States. Whereas that CSA has two major league teams, 13 London teams have played in the Premier League, six during the 2005/06 season.⁵⁹ Using a road map for guidance, we calculated populations for ersatz English CSAs of roughly the same area as Chicago’s. London’s “CSA” has a substantially higher population than the London conurbation, but still only half that of New York City. Again, only two New York City teams now play in the major leagues. Chicago’s CSA has a larger population than our imaginary Manchester-Liverpool CSA, but, as contrasted with Chicago’s two major league clubs, 13 football teams in Lancashire and Cheshire have reached the Premier League, seven during the 2005/06 season.⁶⁰

Table 1 lists, by size over time, British conurbations and cities with teams in the Premier League. The number of teams in the original Football League and the number of teams today give an indication of league growth, with the last four columns indicating when it occurred. All conurbations plus the city of Nottingham have multiple teams with Premier League experience, and other Football League teams in and around nearly all of those places are yet to be promoted that far. The pattern is extremely varied in comparison with North America. A conurbation with over 7 million people hosts 13 past and present Premier League teams, but a town with a population of less than 90,000 also supports one.⁶¹

can add a team from a different nation only if the latter has no domestic league. A violation would result in disqualification from FIFA-sanctioned events such as the World Cup for the national team and the Champions League for club teams. Thus Monaco was permitted to join the French league and Auckland the Australian league. A separate Scottish league formed in 1890, however, so the two major Glaswegian teams, Celtic and Rangers, have been stymied in recent bids to join the Premier League despite a chronic dearth of strong Scottish opponents. Teams from Northern Ireland cannot compete in English, Scottish, or Irish leagues. Cardiff, Swansea, Wrexham, and a few other Welsh teams entered the Football League before the Welsh League began play, and are grandfathered in.

⁵⁹ Those 13 teams include the former Wimbledon, which now plays in Milton Keynes, 55 miles away.

⁶⁰ Accrington, the sole defunct Football League founding member, was included.

⁶¹ Green Bay, Wisconsin, with a 2000 population of 226,778, is by far the least populous North American metropolitan area with a major professional sports team; the next smallest is Edmonton, Alberta, whose population of 937,845 approaches that of Tyne and Wear, the smallest of the seven English conurbations.

TABLE 1
FOOTBALL LEAGUE TEAMS WITH PREMIER LEAGUE EXPERIENCE BY URBAN
AREA, 1888–2006

Rank	2001 Population	Original Teams	1888– 1900	1900– 1915	1915– 1939	1946– 2006
Conurbations						
1 London*	7,172,036			3	6	13
2 West Midlands	2,555,596	3	4	4	4	5
3 Greater Manchester **	2,482,352	1	5	5	5	5
4 West Yorkshire	2,079,217			2	4	3
5 Merseyside	1,362,034	1	2	2	2	2
6 South Yorkshire	1,266,337		2	2	2	3
7 Tyne and Wear	1,075,979		2	2	2	2
Cities						
1 Bristol City	380,615			1		1
3 Cardiff	305,340				1	1
4 Leicester	279,923			1	1	1
5 Nottingham	266,995	1	2	2	2	2
7 Brighton and Hove	247,820					1
11 Stoke-on-Trent	240,643	1	1	1	1	1
12 Swansea	223,293					1
13 Derby	221,716	1	1	1	1	1
14 Southampton	217,478					1
16 Northampton	194,477					1
19 Portsmouth	186,704				1	1
20 Luton	184,390					1
22 Swindon	180,061					1
29 Grimsby (NE Lincolnshire)	157,983			1	1	1
36 Blackpool	142,284				1	1
39 Blackburn w/Darwen ***	137,471	2	3	1	1	1
40 Middlesbrough	134,847			1	1	1
41 Oxford	134,248					1
45 Preston	129,642	1	1	1	1	1
46 Norwich	121,553					1
49 Ipswich	117,074					1
51 Carlisle	100,734					1
52 Burnley	89,541	1	1	1	1	1

* includes the teams in Wimbledon and Watford.

** includes the team in Glossop West End.

*** includes the team in Accrington.

Sources: The Football Leagues http://www.the-english-football-archive.com/football_league.htm; and Office for National Statistics, *Census 2001: First results on population for England and Wales* (2002).

Beginning with the National League's founding and ending with the 1903 Major League Agreement, Table 2 shows, on a year-by-year basis, all teams playing in the National League (denoted by N), the American Association (from 1882 through 1891, denoted by A), the American League (as of 1901, also denoted by A), or that a city had a team in both extant leagues (denoted by 2). Urban areas have been ranked by the 1890 Census; Brooklyn is ranked with New York City throughout these

years. Three noteworthy features are no city with more than two teams until 1903 when the American League placed a team in New York City (see note 24); headlong abandonment of small cities, both features in sharp contrast with English football; and league saturation of the largest markets, comparable to the Football League's expansion north to Tyne and Wear, east to Yorkshire, and south to London.

In Table 3, today's major league cities are ranked at the four census years that correspond as nearly as possible with the Major League Agreement (1900 Census data paired with teams in 1903), the beginning of relocation (1950 paired with 1953), the beginning of expansion (1960 paired with 1961), and today (2000 paired with the present).⁶² At each date, markets without teams at that time are in italics. The process of transfer and expansion that led to teams in an increasing number of major markets is evident. Milwaukee, the smallest market with a team now in Major League Baseball, is nearly 19 times the size of Burnley, the smallest city that has had a team in the Premier League.⁶³ By 1903 teams were in nearly every major market of the northeast quadrant of the United States, but nowhere else. Only three unoccupied markets with populations that exceeded the smallest major league city's would have been available to an entrant in the same region—Baltimore, Milwaukee, and Buffalo (the eighth most populous in 1900 and a perennial bridesmaid in baseball). Before air travel became possible, San Francisco and New Orleans (the twelfth most populous) were too distant to be included. In sum, at the time of the Major League Agreement 16 teams played in ten cities, with the smallest ranked fifteenth (see Figure 1). Those teams remained in precisely the same cities for half a century, but by then the smallest market—Cincinnati—ranked seventeenth, and seven larger ones around the nation had no major league team. Though still tedious, air travel was rapidly improving. Soon the entire nation would comprise a single market for professional baseball. Major League Baseball formulated a plan to meet the coming threat—transfer “surplus” teams from multiteam markets to large cities that had none. In successive years beginning in 1953, one of the two teams in Boston, St. Louis, and Philadelphia moved to Milwaukee, Baltimore, and Kansas City, three sizable cities without major league teams.

⁶² Because several cities that now have major league teams were much smaller in 1900, a few relevant cities are omitted from the table's earlier years.

⁶³ Milwaukee has had a team in the American or National League during 51 of the 130 seasons (39 percent) since the National League began play, and has been home to one World Series Champion (1957); Burnley has had a Premier League team during 51 of the 107 seasons (48 percent) of competition since the Football League was founded, and has been home to two champions (1921 and 1960).

TABLE 2 — continued

	90*	91*	92	93	94	95	96	97	98	99	00	01*	02*	03*
New York	1	N	N	N	N	N	N	N	N	N	N	N	N	2
Brooklyn	1	2	N	N	N	N	N	N	N	N	N	N	N	N
Chicago	2	N	N	N	N	N	N	N	N	N	N	2	2	2
Philadelphia	3	2	2	N	N	N	N	N	N	N	N	2	2	2
St. Louis	4	A	A	N	N	N	N	N	N	N	N	N	2	2
Boston	5	N	2	N	N	N	N	N	N	N	N	2	2	2
Baltimore	6	A	A	N	N	N	N	N	N	N	N	A	A	
Pittsburgh	7	N	N	N	N	N	N	N	N	N	N	N	N	N
San Francisco	8													
Cincinnati	9	N	2	N	N	N	N	N	N	N	N	N	N	N
Cleveland	10	N	N	N	N	N	N	N	N	N	N	A	A	A
Buffalo	11													
New Orleans	12													
Washington	13		A	N	N	N	N	N	N	N		A	A	A
Detroit	14											A	A	A
Milwaukee	15		A									A		
Newark	16													
Minneapolis	17													
Jersey City	18													
Louisville	19	A	A	N	N	N	N	N	N	N				
Rochester	20	A												
St. Paul	21													
Kansas City	22													
Providence	23													
Denver	24													
Indianapolis	25													
Albany	26													
Columbus	27	N	N											
Syracuse	28	A												
Worcester	29													
Toledo	30	A												
Richmond	31													
New Haven	32													
Paterson NJ	33													
Lowell MA	34													
Nashville	35													
Scranton	36													
Fall River MA	37													
Troy	38													
Cambridge	39													
Omaha	40													
Atlanta	41													
Memphis	42													
Camden	43													
Wilmington	44													
Dayton	45													
Grand Rapids	46													
Reading PA	47													
Trenton	48													
Lynn MA	49													
Charleston SC	50													
Hartford	51													

* identifies a year in which there were two leagues.

N = National League.

A = American Association (1882–1891) or American League (as of 1901).

2 = The city had a team in both extant leagues.

Source: Seymour, *Baseball*, vol. 1, pp. 363–64.

For a few years the West Coast remained unworkably distant, but that need not have stymied a major league caliber entrant that confined itself to the region, as the existing leagues continued to confine themselves to the continent's opposite side. The minor Pacific Coast League (PCL) appeared to be on the verge of becoming a third major league, relying solely on the region's population and wealth. The PCL had four teams playing in urban areas that ranked among the nation's seven largest by 1950 and one more ranked nineteenth.⁶⁴ Moreover, all of those markets were growing rapidly, as Table 3 illustrates.

The PCL moved too slowly, being preempted by the 1958 transfer of two of the three New York City franchises to the two largest West Coast markets once long-distance air travel emerged. During the 1950s the smallest major league market had fallen to twenty-second, and air travel between New York and Los Angeles became more convenient than travel by rail between New York City and Chicago had been in 1900.

In 1960, after Major League Baseball had reduced its supply of multi-team cities to Chicago, the insurgent Continental League announced it would place new teams into many of the unoccupied large markets as well as New York City. As a consequence, in 1961 the American and National Leagues initiated a process of expansion that created the first new teams to be admitted to the major leagues since 1901. Once again baseball preempted an upstart, this time by quickly placing expansion teams in three of the Continental League's most attractive intended markets.

By 2000 major league teams were in all of the nation's top 22 markets (and two smaller ones). Except where going head-to-head with the existing leagues, no entrant could have even one team in any market ranking higher than twenty-third. By contrast, at the founding of the American League, its three unopposed franchises were in the seventh, thirteenth, and fifteenth largest markets. During any season, only a few large U.S. markets remain on the outside, hat in hand, hoping wanly for team movement or expansion, almost begging to finance fancy new facilities.⁶⁵

CONCLUSION

By the late nineteenth century, bodies existed to rule sports on both sides of the Atlantic. These bodies faced the potentially crippling externality

⁶⁴ The PCL had two teams in Los Angeles and one each in San Francisco, Oakland, Seattle, Portland, San Diego, and Sacramento, in order of their 1950 populations. Counting Anaheim with Los Angeles (as the Los Angeles CSA does), five of those cities now house six relocated or expansion major league teams.

⁶⁵ For instance, designed as a dual use (football/baseball) stadium, the Louisiana Superdome in New Orleans was completed at public expense 30 years ago. A football team immediately began to play there, but no Major League Baseball team has arrived.

TABLE 3
MAJOR LEAGUE BASEBALL TEAMS BY URBAN AREA, 1900–2000

Rank	Metropolitan Areas				2000 Population
	1900	1950	1960	2000	
1	New York	New York	New York	New York	21,361,797
2	Chicago	Chicago	Los Angeles	Los Angeles	16,373,645
3	Philadelphia	<i>Los Angeles</i>	Chicago	Chicago	9,312,255
4	St. Louis	Philadelphia	Philadelphia	Washington-Balt.	7,538,385
5	Boston	Detroit	Detroit	San Fran.-Oakl.	7,092,596
6	<i>Baltimore</i>	Boston	San Fran.-Oakl.	Philadelphia	5,833,585
7	Cleveland	<i>San Fran.-Oakl.</i>	Boston	Boston	5,715,698
8		Pittsburgh	Pittsburgh	Detroit	5,357,538
9	<i>San Francisco</i>	St. Louis	St. Louis	Dallas-Fort Worth	5,346,119
10	Cincinnati		Washington	Houston	4,815,122
11	Pittsburgh	Cleveland	Cleveland	Miami *	4,452,557
12		Washington	Baltimore	Atlanta	4,548,344
13	Detroit	<i>Baltimore</i>		Seattle-Tacoma	3,604,165
14	<i>Milwaukee</i>	<i>Minneap.-St Paul</i>	<i>Minneap.-St Paul</i>	Minneap.-St Paul	3,271,888
15	Washington			Phoenix *	3,043,878
16		<i>Milwaukee</i>	<i>Houston</i>	Cleveland	2,945,831
17		Cincinnati	Milwaukee	St. Louis	2,754,328
18				San Diego *	2,698,687
19	<i>Minneapolis</i>	<i>Seattle</i>	<i>Seattle</i>	Pittsburgh	2,525,730
20		<i>Kansas City</i>	<i>Dallas</i>	Denver	2,449,054
21		<i>Houston</i>	Cincinnati	Tampa-St. Pete. *	2,395,997
22	<i>Kansas City</i>		Kansas City	Cincinnati	2,050,175
23	<i>St. Paul</i>	<i>Dallas</i>	<i>San Diego</i>		
24		<i>Atlanta</i>	<i>Atlanta</i>		
25	<i>Denver</i>		<i>Miami</i>	Kansas City	1,901,070
26			<i>Denver</i>		
27	<i>Allegheny</i>				
28		<i>Denver</i>			
29					
30					
31			<i>Tampa-St. Pete.</i>	Milwaukee	1,689,572
32		<i>San Diego</i>			
33					
34					
35					
36	<i>Los Angeles</i>				
37		<i>Miami</i>	<i>Phoenix</i>		
38					
39					
40					
41					
42					
43	<i>Atlanta</i>				
44					
45					
46			<i>Fort Worth</i>		
47		<i>Tampa-St. Pete.</i>			
48	<i>Seattle</i>				

Note: Markets without teams are shown in italics.

Source: U.S. Census Bureau. 1900, 1950, and 1960: <http://www.census.gov/population/documentation/twps0027/tab13.txt>; and 2000 population: <http://www.census.gov/population/www/cen2000/phc-t29.html>, table 8 (or from table 3a if city is marked with an asterisk).

that plagues competitive sports. The production of an attractive game is a joint product, and, whereas each team typically bears the marginal cost it incurs in making the game more attractive, the marginal revenues that result from such an improvement are shared. What Dale Mortensen has termed a “mating game” can thus result, with insufficient inputs invested in aggregate to maximize joint profits.⁶⁶ By centralizing decision-making concerning rules of competition and cooperation, the ruling bodies avoided opportunistic behavior and disciplined teams that were in violation. Those bodies gradually evolved during the second half of the nineteenth century. In 1876 the National League revolted against the very ineffective National Association; in 1888 the Football League emerged in cooperation with the very effective Football Association.

That a successful league must be a cartel fails to tell whether the cartel will be open or closed. The number of outside entrepreneurs with a reasonable prospect of entering and the number of inside entrepreneurs facing a reasonable danger of failure may be a more meaningful predictor of behavior than the number of firms that are actually in an industry at any given moment.⁶⁷ North American leagues staunchly control entry and exit; European leagues barely control either, and thus are substantially more contestable.⁶⁸

We have tried to resolve the puzzle of why European leagues became open at almost exactly the same time as North American leagues became closed. The excitement of closely contested games required little travel in the Football League, and compactness provided a way both to limit peripheral players’ wage demands and increase attendance. The potential additional revenues from territorial monopolies could not offset the increased costs. The number, quality, and density of established English teams made a closed cartel untenable, and promotion and relegation offered a way to sort the teams.

North Americans are more numerous than the British, but they are much more thinly spread. That made attracting visiting fans nearly impossible in the nineteenth century, meant that full professionalism arose much earlier than in Europe, and facilitated territorial monopolies, a closed cartel, that reduced demand elasticities. Leaving a few territories unserved at any time, incumbent teams eventually discovered that they could extract substantial entry fees from new franchises, and reduce team costs by extorting stadium construction from political units fearful of losing existing franchises or hopeful of gaining new ones.

⁶⁶ See Mortensen, “Property Rights.”

⁶⁷ Demsetz, “Why Regulate Utilities?”

⁶⁸ The theory of contestable markets originated with Baumol, Panzar, and Willig, *Contestable Markets*. Also see Baumol, “Uprising.”

Our explanation leaves many questions for future work. In particular, why did leagues in other sports adopt the structure of their continent's pioneer league? Learning-by-doing certainly played a role, because debates among teams in other sports demonstrate that the differences between the structures were well understood. But in view of the convergence of the continents' population densities, real wealth, and travel costs, Congressional and court deference to the case of *Federal Baseball Club of Baltimore, Inc. v. National League of Professional Baseball Clubs* must have played a pivotal role. If so, the evolution of sports leagues provides a crystalline example of path dependence. Had air travel arrived a century earlier, it seems unlikely that North American sports leagues would have become closed cartels, thus the world's outliers.

In the small, this article is about the evolution of professional sports leagues. It is indicative of something rather larger, however. Many industrial structures vary noticeably across superficially similar economies, but the causes of those differences are apt to remain obscure until close attention is paid to their historical origins. An industry's economic history matters.

BIBLIOGRAPHY

- Arnold, Tony. "Rich Man, Poor Man: Economic Arrangements in the Football League." In *British Football and Social Change: Getting into Europe*, edited by John Williams and Stephen Wagg, 48–63. London: Leicester University Press, 1991.
- Travellers' Official Railway Guide for the United States, Canada and Mexico*. Edited by W. F. Allen. New York: National Railway Publication Co., December 1899.
- Baade, Robert, and Alan Sanderson. "Field of Fantasies," *Intellectual Ammunition*, <http://www.heartland.org/archives/ia/marapr96/cover.htm> (1996).
- Baumol, William J. "Contestable Markets: An Uprising in the Theory of Industry Structure." *American Economic Review* 72, no. 1 (1982): 1–15.
- Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and the Theory of Industrial Structure*. New York City: Harcourt Brace Jovanovich, 1982.
- Bradshaw's General Railway and Steam Navigation Guide, for Great Britain and Ireland*. London: W. J. Adams & Sons, August 1882.
- British Broadcasting Company. *Champions League Winners Liverpool Have Earned £20.5 m From Their European Run to Top the Tournament's Earnings Table*. <http://news.bbc.co.uk/sport1/hi/football/europe/4092342.stm> (14 June 2005).
- Cain, Louis P., and David D. Haddock. "Measuring Parity: Tying Into the 'Idealized' Standard Deviation." *Journal of Sports Economics*, forthcoming.
- Canes, Michael E. "The Social Benefits of Restrictions on Team Quality." In *Government and the Sports Business*, edited by Roger G. Noll, 81–114. Washington, DC: The Brookings Institution, 1974.
- Chandler, Alfred D., Jr. *Scale and Scope: The Dynamics of Industrial Capitalism*.

- Cambridge, MA: The Belknap Press of Harvard University Press, 1990.
- Demsetz, Harold. "Why Regulate Utilities?" *Journal of Law and Economics* 11, no. 1 (1968): 55–65.
- Eckard, E. Woodrow, "Team Promotion in Early Major League Baseball and the Origin of the Closed Sports League." *Explorations in Economic History* 42, no. 1 (2005): 122–52.
- Football Grounds From the Air: Then & Now*. Hersham: Ian Allen Publishing, 3rd edition, 2004.
- Goldstein, Warren. *Playing for Keeps: A History of Early Baseball*. Ithaca, NY: Cornell University Press, 1989.
- Harris, Harold Arthur. *Sport in Britain: Its Origins and Development*. London: Stanley Paul, 1975.
- Hauptert, Michael, and Kenneth Winter, "Pay Ball: Estimating the Profitability of the New York Yankees, 1915–1937." University of Wisconsin, LaCrosse Working Paper, 2004.
- Kirsch, George B. *The Creation of American Team Sports: Baseball and Cricket, 1838–72*. Urbana: University of Illinois Press, 1989.
- Knautz, Robert, editor. "Franchise Free Agency." *Policy Spotlight*, <http://www.free-market.net/features/spotlight/9702b.html> (1997).
- Leeds, Michael, and Peter von Allmen. *The Economics of Sports*. Boston: Addison Wesley, 2002.
- Mortensen, Dale T. "Property Rights and Efficiency in Mating, Racing, and Related Games." *American Economic Review* 72, no. 5 (1982): 968–79.
- Nichols, Dan. *Major League Franchise Information*, <http://www.kpris.com/Baseball/bb-franchiseInfo.html>.
- Noll, Roger G. "The Economics of Promotion and Relegation in Sports Leagues." *Journal of Sports Economics* 3, no. 2 (2002): 169–203.
- Okner, Benjamin A. "Subsidies of Stadiums and Arenas." In *Government and the Sports Business*, edited by Roger G. Noll, 325–48. Washington, DC: The Brookings Institution, 1974.
- Poitras, Marc, and Lawrence Hadley. "Do New Major League Ballparks Pay for Themselves?" *Journal of Business* 79, no. 5 (2006): forthcoming.
- Quirk, James, and Rodney Fort. *Pay Dirt: The Business of Professional Team Sports*. Princeton, NJ: Princeton University Press, 1992.
- Rader, Benjamin G. *Baseball: A History of America's Game*, second edition. Urbana: University of Illinois Press, 2002.
- Robinson, Michael, editor. *Non-League Football Tables 1889–2004*. Cleethorpes, Lincolnshire: Soccer Books Limited, 2004.
- Ross, Stephen F., and Stefan Szymanski. "The Law & Economics of Optimal Sports League Design." University of Illinois College of Law, Illinois Public Law and Legal Theory Research Papers Series, Working Paper No. 03–14, 30 September 2003.
- Seymour, Harold. *Baseball: The Early Years*, 3 volumes. New York: Oxford University Press, 1960.
- Stigler, George. *The Organization of Industry*. Homewood, IL: Richard D. Irwin, 1968.
- Surdam, David. "The American 'Not-So-Socialist' League in the Postwar Era: The Limitations of Gate Sharing in Reducing Revenue Disparity in Baseball." *Journal of Sports Economics* 3, no. 3 (2002): 264–90.
- Szymanski, Stefan. "The Economic Design of Sporting Contests." *Journal of Eco-*

- conomic Literature* 41, no. 4 (2003): 1137–87.
- Szymanski, Stefan, and Andrew Zimbalist. *National Pastime: How Americans Play Baseball and the Rest of the World Plays Soccer*. Washington, DC: Brookings Institution Press, 2005.
- Taylor, Matthew, and John Coyle. “The Election of Clubs to the Football League, 1888–1939.” *The Sports Historian* 19, no. 2 (1999): 1–24.
- Vamplew, Wray. *Pay Up and Play the Game: Professional Sport in Britain, 1875–1914*. Cambridge: Cambridge University Press, 1988.
- White, G. Edward. *Creating the National Pastime: Baseball Transforms Itself, 1903–1953*. Princeton, NJ: Princeton University Press, 1996.
- Young, Percy M. *A History of British Football*. London: Stanley Paul, 1968.

CASES CITED

- American League Baseball Club v. Chase*, 149 N.Y.Supp. 6 (1914).
- Federal Baseball Club of Baltimore, Inc. v. National League of Professional Baseball Clubs, et al.*, 259 U.S. 200 (1922), affirming 269 Fed. 681 (App. D.C. 1921), which reversed the trial court.