Stanford Center for International Development

Working Paper No. 180

The Organization of Sports Leagues

by

Roger G. Noll*

August 2003

Stanford University
John A. and Cynthia Fry Gunn Building
366 Galvez Street | Stanford, CA | 94305-6015

* Department of Economics, Stanford University

THE ORGANIZATION OF SPORTS LEAGUES

by Roger G. Noll

Abstract: Sports leagues have been organized in many different ways. This essay examines the incentive structure and efficiency of different forms of league organization, including the methods for scheduling games, admitting new members, and making operational decisions. This article also compares operations and outcomes in Europe and North America, and concludes that the European system of promotion and relegation is superior to the closed structure of American leagues, and that the American system of multiple parallel leagues to determine qualifications and seeding in a post-season tournament is efficiency enhancing. The article also discusses the optimal size and number of leagues, and concludes that both the European and American systems produce too few major league teams, largely because they have permitted major leagues to be monopolies.

Forthcoming, *Oxford Review of Economic Policy*

THE ORGANIZATION OF SPORTS LEAGUES

by Roger G. Noll*

Team sports require coordination among contesting teams because the main product, a game, involves at least two distinct entities. Teams must agree on the rules of the game, the time and venue at which it will be played, the identity of the officials who will enforce the rules and keep the score, the responsibility for marketing the contest if one objective is to collect revenue, and the procedures for dividing the revenues and costs of the contest. These decisions sometimes are made between contesting teams on a game-by-game basis. But if several teams regularly play each other, this approach is highly inefficient, and provides an incentive to create a league – a group of teams that schedule games and develop other policies and rules for the purpose of determining a champion (which among them is strongest).[1]

Team sports almost always are organized into leagues. Despite the simplicity of the concept of a league, the structure of sports leagues varies enormously. Teams in a sport must make at least five types of decisions about league structure:

* format – the method for scheduling matches to determine the champion;
* hierarchy – the relationships between leagues of lesser and greater quality;
* multiplicity – the number of leagues at the same level of the hierarchy;
* membership – the conditions under which a team enters and exits a league; and
* governance – the methods for deciding and enforcing league rules and policies.

A league also normally makes other decisions, such as developing playing rules and controlling aspects of the economic behavior of its members.

This article explores the economics of the organizational form of leagues. The choice of organizational form affects the demand for a sport, the cost of scheduling games, and the extent of competition among teams for fans and their most important inputs, players, coaches and stadiums. The organization of leagues is interesting for two reasons. First, the incentive effects of league rules, and hence the efficiency of league operations, are subtle, and frequently misunderstood by fans, journalists and even team owners. Second, the most prestigious professional leagues, called major, premier, or first division leagues, are almost always monopolies. Thus, teams have a strong incentive to organize leagues in a fashion that reduces the extent of horizontal competition among them in both input and output markets. An important issue in the organization of team sports is to distinguish between rules and policies that improve efficiency from those that reduce efficiency by reducing competition among teams.

A monopoly major league controls the entry of teams into the top echelon of the sport. One must then ask whether a monopoly league's procedures for admitting new members produce the optimal number and geographical distribution of teams. Two elements of this analysis are the optimal number of teams in a city and major leagues in a sport. Specifically, are teams in a local market and leagues at a given quality level "natural monopolies" in that competition is likely to be unstable or otherwise to reduce the net social benefits of a sport? An additional issue is how leagues change the cities in which teams play. Some leagues have permanent, fixed membership and change the cities in the league through expansion and team relocation, while other leagues regularly change their membership by replacing the weakest teams with the best teams in lesser leagues, as is common in

European team sports. Which is more efficient?

The purpose of this article is to analyze these issues. The article begins by discussing the choices available for structuring leagues. The discussion then focuses on the optimal size of leagues, and whether teams and leagues are natural monopolies. A major theme throughout this article is the comparison between "American" of "European" organizational forms. The former refers to leagues with fixed membership and territorial exclusivity, and the latter refers to the "promotion and relegation" system.

VARIETIES OF ORGANIZATIONAL FORMS

A team sport can choose among a variety of organizational forms. The first decision is to create a league. Teams form leagues in part because players, coaches and owners enjoy contesting for a championship, but the primary factor determining the form of professional leagues is its financial consequence. Leagues create the opportunity to market a game as both the contest itself and one of a series that leads to a championship. Because the quest for a championship generates fan interest, league matches leading to a championship are covered more extensively by the media, which thereby provide free promotion. Leagues also reduce transactions costs by enabling teams to coordinate scheduling, rather than relying on a series of bilateral agreements.

Studies of the demand for team sports, such as the article in this issue by Jeffrey Borland and Robert Macdonald, show that attendance depends, among other things, on the significance of the contest with respect to a championship.[2] A team that has dropped out of contention for a championship generally will draw poorly, but it is likely to sell more tickets if it is playing a team that is at or near the top of the standings than if it is playing another weak team, even though the outcome of the latter game is more uncertain.

Once teams decide to adopt a league structure, the members face additional decisions about how the league will be organized and governed. Leagues are not completely free to determine their structure. Teams and leagues are also constrained by government policies, such as antitrust and labor laws, and by non-governmental governing bodies in each sport. The role of policy is discussed throughout this article, and the role of governing bodies is discussed in the section on governance. The analysis to follow focuses on autonomous leagues as the unit of analysis, although it applies to governing bodies when they organize leagues. For example, the International Federation of Football Associations (FIFA) organizes soccer's World Cup, which is contested by national All-Star teams, and the Union of European Football Associations (UEFA), a subordinate body to FIFA, organizes the European Champions League, involving the top professional teams during the previous season from the major national leagues in Europe.

Format

A league can adopt two types of schedules: a round-robin or an elimination tournament.

In a round-robin, the league creates a schedule of games for a championship season for each team, and every team plays a predetermined number of games against other league members. The champion is determined by aggregating the results of all matches. The schedule can be "balanced" (all teams play all others an equal number of matches) or "unbalanced" (each team plays some teams more than others), and an unbalanced schedule can be "even" (every team plays the same number of league games) or "uneven." Balanced schedules are the norm in Europe, but not in North America, and while most unbalanced league schedules are even, a few are not.

Sports purists regard a balanced schedule as superior because it produces a final league standing that is most likely to rfeflect the actual rank-ordering of teams by quality. Regardless of the aesthetic demerits of an unbalanced schedule, it can enhance demand. Unbalanced schedules increase the uncertainty of match outcomes because the prior records of the contesting teams are less comparable. To the extent that outcome uncertainty enhances demand, unbalanced schedules increase revenues from both the round-robin schedule and the playoffs.

In an elimination tournament, teams are dropped from the schedule after losing a certain number of games, usually one (single elimination) or two (double elimination). Soccer frequently uses a variant of the single-elimination tournament in which two teams play a game in each home stadium with the winner determined by aggregate score. Tournaments can be "seeded" (stronger teams play weaker ones early on, thereby maximizing the likelihood that the strongest teams will survive to the end of the tournament) or "drawn" (matches are determined by random draw, in which case two of the best teams can meet in an early round). A tournament is unbalanced if some teams must play more games than others to reach the final match. In England, the Football Association (FA) cup, which now has over 500 entrants, is seeded and unbalanced in that weaker teams play several games to determine which among them is allowed to compete against teams from the Football League.[3]

Many leagues schedule both a tournament and a round-robin. In all North American professional leagues, some European basketball leagues and the French Rugby Union, a round-robin determines which teams qualify for a post-season tournament and the seeding in a single-elimination playoff. In England, the two major tournaments, the FA Cup (open to all teams in FA sanctioned leagues) and the Football League (FL) Cup, open to teams in the four professional leagues, and several minor tournaments with more restricted eligibility, are carried on simultaneously with league schedules. Likewise, the European Champions League runs in parallel with league schedules. The Champions League and World Cup are organized in the American style: four-team round-robins determine qualifications and seeding in an elimination tournament.

The economic advantage of tournaments over round-robins is that tournaments substantially increase the importance of each match, and thereby create more intense demand for each game. The disadvantage of tournaments is that elimination leads to few games for most teams, which sacrifices capitalizing on the demand for matches that may not matter in determining a championship. As a practical matter, virtually all leagues organize both round-robin and elimination schedules, implying that they can capture the benefits of both systems from dual championships; however, theoretically, operating both formats, either simultaneously as in England or sequentially as in North America, has an ambiguous financial effect. The benefit is that it creates two different championships that can enhance demand, especially if the luck of outcomes causes at least some different teams to contend for each championship. The cost is that a second championship can detract from the value of the first, thereby leading to lower demand for each than would be the case if either were to be the only format in place.

Another problem with multiple formats is that weak teams in strong leagues are likely to prefer a round-robin. Because an elimination tournament reduces the number of games that could be scheduled in a round robin, a weak team that has a high probability of early elimination expects to experience a double loss: less demand because the round-robin championship is less important, and less revenue because it plays fewer games. This effect is unlikely to be important for minor-league teams that engage in all-sport tournaments such as the FA or FL cups, for their leagues expect that all of their teams will be eliminated early and so schedule more games in the round-robin. For example, lower leagues in England, ineligible for European championships and unlikely to survive long in major cups, have more members and schedule more games than the Premier League and Division One.

Hierarchy

Most team sports are organized in a hierarchy of leagues. The top of the hierarchy contains the major league, premier league, or first division, followed by minor leagues or lower divisions. In English football, the hierarchy is extremely deep, with ten levels of the Football Association (including the upper and lower divisions of some lower minor leagues). In the United States, universities function as an important substitute for lower minor leagues, so that the hierarchy of professional teams is shallower than in European football. American baseball has the most levels, with five classifications of professional leagues (Major, AAA, AA, A, Rookie).[4] American soccer and hockey have two minor levels, and basketball and football have one.

American intercollegiate sports are organized hierarchically as divisions of the National Collegiate Athletic Association (NCAA), which has four divisions for American football and three for other sports. The National Association for Intercollegiate Athletics provides another division below NCAA Division III, except in basketball where the NAIA has two divisions. If one includes college leagues, the number of levels in America are similar to those in England.

Sports became organized into hierarchies of leagues is response to the nature of demand. In America, but almost nowhere else, identification with local universities creates a distinct demand for college sports that has relatively little competitive overlap with professional sports. For professional sports, demand in some localities is not sufficient to support a team that would attempt to compete with teams that enjoy much greater intrinsic demand. For example, suppose that all 92 teams in the English Football League joined the Premier League. Current lower division squads would be forced either to spend much more on players and coaches, or to have no hope of being in contention for the championship, or even of winning more than a handful of games. Because the chance of winning a 92-team championship is so low, most teams would not try, and so would be over-matched in many games. Whereas home games against Premier League powerhouses might draw well, the size of the league would preclude scheduling more than a few such games.[5] For other matches, demand would suffer because the games would not be consequential in determining the championship. Thus, teams that would have contested for a Division 2 or 3 championship would suffer a general decline in support. Likewise, home games of powerhouses against weak squads would be less well attended than the games that they replaced because fans would have less interest in mismatches. The number of viable teams likely would be larger in a hierarchy as some teams that could not survive in a higher league would survive in a lower league with its demand-enhancing championships and lower costs. Consequently, the sport as a whole would do better financially by organizing into a hierarchy of leagues.

Hierarchies of leagues provide other benefits to weaker teams. A hierarchical structure can provide the benefits of occasional matches against top teams by organizing exhibitions ("friendlies") and tournaments, such as the FA Cup and FL Cup. In addition, teams in a lower leagues can serve as training grounds for young players who are not yet skilled enough to compete at the highest level. If these teams identify promising players and sign them to long-term contracts, after the player develops they can sell their contracts to teams in a higher league. These transfers are in the interests of both teams because a skilled player is worth more in a higher league, so that a player not only derives personal satisfaction from playing in a better league, but earns higher wages, especially after the initial contract expires.

Multiplicity

Sports vary in the extent to which each level in the hierarchy is divided into multiple leagues or divisions. The term "multiplicity" is better than "competition" since multiple leagues at the same hierarchical level may or may not compete, depending on the leagues' membership rules and other operating procedures. Multiple leagues at the same hierarchical level frequently are jointly managed, being primarily a marketing device that is used to enhance demand by creating multiple championships and increasing uncertainty in qualifications for a post-season tournament.

In Europe, each nation has a single league at the top professional level. American professional sports have multiple leagues or divisions at the top of the hierarchy. In baseball, both North America (American, National) and Japan (Central, Pacific) have two major leagues. The American baseball leagues are each divided into three divisions. In North America basketball, football and hockey have a single major league, but each has two distinct conferences and then divisions within each. Even Major League Soccer in the U.S. is divided into two divisions. In all North American leagues, including soccer, divisional standings determine qualifications and seeding for a post-season tournament. Teams have some inter-league or inter-conference matches that count in the standings, but the schedules are unbalanced.

For lower leagues, the pattern is mixed. In England, one must descend to the sixth level of the football hierarchy – below the Football Conference – to find multiple leagues at the same level. In

North American baseball, all levels in the hierarchy have at least two leagues, starting with the National and American at the major league level and descending to several leagues at the A level and two at the rookie level.

In American football, the second level of the hierarchy has three unusual leagues. The Canadian Football League (CFL) is the only independent minor professional football league in North America. The CFL plays on a larger field and has other rule differences that cause differences in strategy and the optimal mix of player skills, much like the differences between Rugby League and Rugby Union. Nevertheless, despite these differences many CFL players, after proving their skills, move on to the NFL. Although historically having only Canadian members, in the 1990s the CFL attempted to enter the United States, but the experiment failed and the CFL is again a purely Canadian league. The other two minor football leagues are owned by the NFL. The Arena Football League plays indoors on a small field and uses fewer players, making it less than ideal as a training ground for the NFL, and the World League plays in Europe. Both play during the late winter and spring, whereas the NFL seasons runs from the late summer to early winter, and some players who are back-ups in the NFL also play in one of these leagues.

Basketball has several minor leagues, and no formal mechanism for classifying them. Five leagues operated in 2003.[6] In hockey, the number of leagues is shrinking. In 2000, hockey had two leagues below the major National Hockey League, and five leagues at the next level.[7] At the end of the 2002-3 season, the field shrank to one at the second level and three at the third level.[8] Most teams from the defunct leagues remain, having been merged into the other leagues, so the contraction was more a reduction in leagues than teams.

American intercollegiate teams are organized into many leagues in each level of the hierarchy. The top football category, Division IA, currently has 114 teams divided into eleven leagues (plus a few independents). The six conferences that dominate the sport sponsor the "Bowl Championship Series," or four post-season games involving their league champions plus two other highly rated teams. One BCS game matches the two most highly rated teams to determine the national champion.[9] The lower football divisions and other intercollegiate team sports in all divisions have national championship post-season tournaments involving league champions plus other highly-rated teams.

Sports differ in the extent to which teams can belong to multiple leagues. Whereas exclusivity – a team can belong to only one league – is common, some multiple memberships are permitted. The most important example of multiple memberships is the European Champions League. In 2003, three Italian teams simultaneously were battling for the Serie A championship while playing in the semi-finals of the European Champions League.

Multiple memberships, like multiple championships from different formats, have an ambiguous effect on economic welfare. As with multiple formats, multiple memberships can reduce the value of each championship. Moreover, simultaneous participation in several leagues could lead a team that, through bad luck, had dropped out of contention in one league to focus all of its energies on another, thereby reducing the quality of play in (and hence the demand for) its matches in the first. For example, as the season nears an end, a team may rest its star players in games in one league in order to make them fresher for matches in the other. Finally, the opportunities for multiple memberships are not likely to be the same for all teams, which can cause problems for the more inclusive league. For example, only a few teams are granted the opportunity to play in the European Champions League; the remaining teams in the top European national leagues have much lower revenues because they have no comparable opportunities. This opportunity gap leads to a larger gap in team quality, reducing the competitiveness of (and hence the demand for) league matches. As a result, the rising popularity of the Champions League eventually could undermine the viability of national leagues (Hoehn and Szymanski, 1999).

For these reasons, multiple memberships are less common and more controversial than multiple formats, and explain why, for a while, England prohibited its teams from participating in the European Championships. Leagues normally require exclusive membership. Acquiescence by leagues and national governing bodies to the European Champions League is a response to the strong incentives acting upon the best professional clubs and television broadcasters to form an international league that features the perennial powerhouses of the national leagues.[10] The alternative to acquiescence plausibly is not to ban the Champions League, but for the best teams to withdraw from their national leagues to form a European super-league that, in turn, would undermine the national governing bodies of soccer.

Membership

An important element of league organization is the procedure for determining its members. All leagues can expel a team if it fails to attend matches, does not field a team of appropriate quality, violates leagues rules in acquiring and paying players, plays in a substandard facility, or brings dishonor to the sport. While expulsions are rare, they are not unknown.

Contraction and Expansion

In recent years, several leagues have debated contraction, whereby a league shrinks by expelling members that do not want to withdraw and that have not violated any league rules. Three examples are: (1) the shrinkage of the Premier League from 22 to 20 members; (2) the announcement by ten of the twelve members of the Scottish Premier League to resign and then to form a new Scottish first division that would exclude the "Old Firm," Celtics and Rangers, because the latter were too dominant; and (3) the decision, now postponed until at least 2007, by Major League Baseball to contract two of the financially weakest teams.

All leagues also have changed their size through expansion (adding teams). For example, English professional football grew from twelve to eighty-eight teams in the three decades after the founding of Division One through a process of adding teams and lower leagues. Nevertheless, in the past few decades, substantially changing the number of major league teams has been common only in North America.

Since 1951 the four divisions of the English Football League have had the same number of members, 92. In 1953 in the United States, major league baseball had sixteen teams, football had twelve, basketball had eight, and hockey had six. (Soccer had none because there was no major league.) By 2003, major league membership had grown to 32 in football, 30 in both baseball and hockey, and 29 in basketball, with teams being added at an average rate of more than one per year. Most of this growth came through expansion, although all sports except baseball also grew by merging with competing leagues (accounting for thirteen teams in football and four each in basketball and hockey).[11]

Relocation

Leagues also have procedures for governing the movement of the home location of teams. In 2002 the Charlotte Hornets of the National Basketball Association moved to New Orleans, and in

2003 the British football team in Wimbledon moved to Milton Keynes. In 1996 the Houston Oilers of the U.S. National Football League announced their move to Nashville to become the Tennessee Titans, stopping in Memphis for a few years while their new stadium was under construction. These moves all required league approval. Although approval was granted in all cases, sometimes proposed moves are highly controversial, and occasionally approval is denied.

Wimbledon's move was hotly contested, despite Wimbledon's decades of low attendance. Despite over a decade in the Premier league, Wimbledon's attendance frequently was exceeded by many Division I clubs. For example, in 1999-2000, Wimbledon finished 18th in the Premier League, but 34th in attendance, being outdrawn by 14 Division 1 teams and one Division 2 club. After being relegated to Division 1, Wimbledon's rank in attendance fell to 45th in 2000-01 and 56th in 2001-2, despite above average finishes in Division 1. In 2001-02, Wimbledon's attendance was next to last in Division 1, and bested by ten clubs in Division 2 and three in Division 3. Clearly the Wimbledon district of London has proven to be a poor location for a high-quality English football squad, so that the heated opposition to Wimbledon's relocation represents the triumph of tradition over rationality. Fortunately for the team and for football fans in Milton Keynes, rationality eventually triumphed.

The United States has far more experience with team relocation, and several proposed movements have been rejected. In the early 1990s, the San Francisco Giants sought approval to move to Tampa, but Major League Baseball rejected the proposal. Connie Mack, the grandson of one of the greatest baseball executives in history and also Senator from Florida, objected strongly to baseball's decision, and submitted legislation to remove baseball's antitrust exemption. Baseball responded by expanding to both Miami and Tampa, but both teams have drawn poorly and were mentioned as candidates for contraction in 2002. Meanwhile, the Giants have become one of the most successful teams, financially and athletically.

Likewise, the NFL's Oakland Raiders sought and were denied approval to move to Los Angeles in the early 1980s, but responded by suing the NFL for violating antitrust law and moving anyway in 1982. The Raiders won the suit, but in 1995 moved back to Oakland as Los Angeles broken its promise to build a new stadium. The outcome of the Raiders' lawsuit set an important precedent, for it requires leagues to set forth clear standards for denying a team's request to move and to adopt procedures for reviewing a proposed relocation that avoid the temptation other teams might have to weaken a competitor financially by denying it a more lucrative market. As a result, team relocation in the U.S. has become more difficult to block since the Raiders' case.

Closed Versus Open Leagues

The most important choice regarding league membership is whether to adopt promotion and relegation. Since creating a lower professional league in 1892, English football has had open leagues: that is, the best teams from a lower league are promoted to the next highest league, while the weakest teams in the latter are demoted to the former. Most European leagues follow the same practice, but in the United States major leagues are closed, having a fixed membership that can only be changed by formally voting to expand. Moreover, American professional leagues set extremely high expansion fees – in the hundreds of millions of dollars – for entering teams. These fees are monopoly prices, reflecting an underlying scarcity in teams that is maintained by monopoly leagues to maximize their income from expansion and the market value of existing teams when they are sold.

As the English Football League added divisions, it continued promotion and relegation between lower minor leagues; however, regular promotion and regulation involving Division 3 and the Football Conference is a recent phenomenon. Originally, teams outside the Football League petitioned for admission to Division 3, and League members then voted whether to admit them and to demote others. In most years, teams were neither relegated from the Football League nor promoted from the Football Conference, so that collectively the four divisions of the FL were *de facto* a closed league.

Beginning in the late 1980s, promotion and relegation between Division 3 and the Football Conference occurred in most years, and recently it was made automatic if the qualifying Football Conference team satisfies Football League requirements for a playing venue. Beginning in 2003, the number of teams promoted and relegated between Division 3 and the Conference was increased from one to two.

An important difference between fixed-membership and promotion-relegation leagues is that only the former can effectively prevent local competition among teams in a league by creating exclusive territorial rights. In the United States, each team has a well-defined "home territory," which usually is an entire metropolitan area. No team can stage matches or even broadcast games within another team's home area without first obtaining permission. This rule protects teams against local competition.

In a few cases, approval has been given, but only after substantial compensation. For example, the New York Mets baseball team paid the New York Yankees $10,000,000 for the right to share New York City. Similar deals were arranged to enable the California Angles to play in Los Angeles and the Oakland A's to play in the San Francisco-Oakland metropolitan area. An interesting feature of territorial rights is manifest in San Francisco. The largest city in the metropolitan area is San Jose, which is part of Santa Clara County and its Silicon Valley, one of the wealthiest communities in the

United States. When the Raiders moved back to Oakland in 1995, the stadium that the A's and Raiders shared was renovated to make it better for football, but much worse for baseball. Santa Clara County officials then sought to induce the A's to relocate, but ran up against baseball's territorial rights. Santa Clara County is in the Giants' territory, despite the fact that the proposed site for the new baseball stadium was forty miles south of the Giants stadium in San Francisco.

Territorial rights are not consistent with the principle under promotion and relegation that a lowly amateur squad, with luck and money, can ascend the hierarchy of leagues to reach the Premiership – as once did Wimbledon, despite the presence of FL teams in nearby Brentford, Chelsea and Fulham. Amateur teams are organized by neighborhood, and even small cities have many. Only by excluding even the lowliest teams from sharing a metro area or blocking these teams from promotion could a Premier League team be sure that it would never face competition from another squad in an adjacent neighborhood. One can not imagine the Giants-A's-San Jose circumstance arising in Europe, where it is unimaginable that a team forty miles away could block another team from playing in a community as populous and wealthy as San Jose.

If the Football League had adopted territorial rights like those enjoyed by U.S. teams, most likely the FL would have far fewer teams. During the 1990s, 14 teams from the London metro area (11 within the London city limits) were members of the FL. In addition, Birmingham, Bristol, Liverpool, Manchester, Nottingham, Sheffield and Stoke all had two FL teams in the same city, and several of these metro areas contain still more teams. With exclusive territorial rights, the Football League could have reached its current size only if fifteen to twenty small cities and towns that now lack FL clubs nevertheless could support one. This outcome seems highly unlikely, given the meager support for many Division 3 teams in small communities, and the even weaker support for nearly all teams in the Football Conference and lesser leagues. For example, in 2002-03, only nine of the 92 Football League teams averaged fewer than 3,500 fans per game, but only three Football Conference teams averaged more than 2,110, the average attendance for the poorest-drawing Division 3 team, Macclesfield. Indeed, only 21 teams in all lower leagues averaged more than 1,000 fans per game. Thus, there are few plausible candidates in the Football Conference to replace the multiple teams in the same cities.

Optimal League Size

The permanence of multiple teams in many English cities, compared to the scarcity of multi-team cities in the U.S., raises the question of the optimal size of a sports league. This question clearly is closely connected to the issues of format, multiplicity and hierarchy as well as the mechanism for determining membership. The issue has four distinct components.

The first is whether a team in a locality is a natural monopoly. If teams are natural monopolies, then regardless of the excess profits generated by the first team, multiple teams in one locality can not be successful. This issue is explored in the next section. Of course, even if only one team can survive in a market, entry restrictions still are not economically desirable. In all industries businesses fail because they are not efficient in satisfying demand. Entry restrictions protect an inefficient businesses from replacement by a more efficient firm. Even in a natural monopoly, competition *for* the market, while not as significant a force for efficiency as competition *in* the market, can improve consumer welfare.

American baseball history provides a useful example. In 1899, the American League entered to compete against the National League, and located teams in direct competition with the National League in Boston, Chicago, New York, Philadelphia and St. Louis. In four cities (Boston, Chicago, New York and Philadelphia), the entrants became the more successful team, and the Yankees and A's jointly dominated baseball during the 1920s and 1930s. In Boston, the National League team eventually fared so poorly that it was driven from the city in the 1950s, while two more teams left New York because the city would not provide stadiums that would put them on equal footing with the Yankees. Although New York, Boston and Philadelphia are not likely to be local natural monopolies, even if they were, competition from the American League was valuable to consumers because it enabled better managed teams to prevail.

The second issue is the welfare economics of a sports team: whether its social benefits exceed its social costs. This issue addresses whether a team makes a net contribution to social welfare. Team sports have two features that assure that the socially optimal number of teams is likely to be far larger than the number of teams that are financially viable: external benefits and player rents.

A team creates external benefits (see Noll and Zimbalist, Chapter Two, for a more complete discussion). Print and broadcast firms derive profits and readers derive consumer surplus from media coverage of team sports, yet they do not pay teams for this coverage. Likewise, some benefits of a team are manifest as conversations among fans about sports. As with media coverage, fans who assemble around the water cooler to discuss yesterday's match and speculate about the outcome of the next pay no royalties to the teams for discussing them.

The welfare of a team and its fans can be enhanced by the presence of a nearby team. Teams in adjacent areas frequently form a natural rivalry, whereby the demand for games with each other is greater, all else equal, than for games with other teams. But adding teams also creates an external cost. Adding a team reduces the probability that each team will win or be in contention for a championship. Because the demand for a team is affected by its position in the championship race, all else equal a greater number of teams causes a lower average demand per team. This factor is likely to cause existing teams to resist expansion even if expansion teams are profitable (Szymanski, forthcoming).

Teams also generate large social benefits because player salaries are mostly economic rents, not true social costs. Professional athletes at the highest level typically are paid wages that exceed the earnings that they could command in their next best occupation. Thus, part of the social surplus of sports is captured by athletes in the form of earnings that exceed the pay that would be necessary to induce their participation in sports. A welfare analysis of American baseball reveals that even the weakest baseball teams generate tens of millions of dollars of net surplus annually, while the best teams generate surplus in the hundreds of millions (Noll, forthcoming). This finding is virtually certain to be qualitatively similar in other team sports.

The third issue in ascertaining the optimal size of a league is to determine how many teams are financially viable, which means determining the number of markets that can support at least one team. The answer to this question depends on the rules of the league and government policies. If teams in financially weaker markets receive extra income from revenue sharing, subsidies and monopolized output markets, and enjoy lower costs from monopsonized input markets, the number of viable teams is greater. As argued above, an open league (no territorial restrictions) can have multiple teams in the best markets, so that all else equal, one would expect promotion and relegation leagues to have more potentially viable teams than closed leagues.

Unfortunately, promotion and relegation systems also reduce the incentive of existing leagues to add members. Since the 1950s, an important factor motivating expansion in the U.S. has been the threat that new leagues would emerge (Quirk and Fort, 1992). New leagues have an incentive to enter because American closed leagues place too few teams in the best markets, and leave other cities with unoccupied viable markets due to the desire to set monopoly entry fees for expansion teams. By contrast, the process of promotion and relegation allows the location of teams to follow demand without the painful and controversial process of relocation. This process allows the composition of league membership to adjust to occupy the best markets, thereby leaving only less attractive potential locations for an entering league. Hence, a league that practices promotion and relegation, because it is less likely to be threatened by entry of a new league, will be less inclined to expand.

The fourth issue that determines the optimal scale of a team sport is the supply elasticity of high-quality players. A common view among fans and sports journalists is that the number of high-quality athletes is the binding constraint on the number of teams, but this conventional wisdom is almost certainly false. In all major sports, the number of teams has expanded less rapidly than population growth for decades, so that the fraction of prime-age adults who are professional team sport athletes is declining. Unless human beings are suffering declining athletic ability, the number of athletes who could play adequately at the professional level is larger than the number of positions available.

In addition, the extent to which the supply of players limits the number of teams depends on league rules about roster size. In North America, all team sports have roster limits that severely bind a team's flexibility in making substitutions or replacing an injured player. In English soccer, teams historically have not had roster limits and so have employed more players in relation to the number who actually see action in a game. Adopting roster limits spreads quality athletes more evenly among teams, and increases the number of viable teams.

Finally, for one nation the supply of professional athletes is essentially unlimited in sports that are played in many countries. Indeed, only American football is so geographically limited that the supply of athletes as seen by the NFL is roughly the worldwide supply of players. In soccer, basketball, baseball, hockey, cricket and rugby, many nations produce quality athletes, so the supply for a single national league is likely to be highly elastic.

The upshot of this section is that the number of teams at the top of the professional hierarchy almost certainly is too small. In closed leagues, the number is too small primarily because big cities have too few teams and because leagues monopolize franchises to maximize their value. In open leagues, the number is too small because existing teams have a strong incentive to stop expanding long before the number of viable members is exhausted and because leagues face no serious threat of entry.

Governance

The two fundamental issues concerning governance of sports leagues pertain to the ownership structure of leagues and teams, and the extent to which an independent authority plays some role in the operations of the league. This section discusses these issues.

Joint Ventures and Single Entities.

Most professional leagues are organized in roughly the same fashion. Member teams are independently owned and managed, and collectively create the league as a joint venture for coordinating their league activities. A commissioner or president is the chief operating officer of the league, while team owners constitute the board of directors. The chief executive may have considerable expressed authority in the league's rules and policies, but true power is in the hands of team owners. Typically the chief executive serves at the pleasure of a majority of the teams, and can be removed without cause if a majority so desire. As a result, the league chief executive can exercise considerable authority in disciplining the bad behavior of one owner or player, but is not likely to survive in office if a decision is made that harms most teams.

A few leagues have been organized as "single entities," with true power centralized in the league office. In these leagues, teams are not independent organizations, but are operating divisions of the league. These leagues also have a chief executive and a board of directors, with the latter being the major investors in the league. In its pure form, team operators in single-entity leagues serve at the pleasure of the league, and can be removed by either the chief executive or the board, depending on the league's rules. Thus, team operators are roughly equivalent to general managers, not owners, in joint-venture leagues.

The most important distinction between joint-venture and single-entity leagues is the degree to which they centralize decisions about resource allocation and league membership. In theory, a single-entity league can achieve an optimal allocation of players and other inputs among teams by simply allocating inputs (assigning players, coaches and managers, and arranging for stadiums, for each team), or by exercising firm budget controls and manipulating the incentives of managers. This feature of a single entity is a practical advantage only under three circumstances. First, fans must accept the legitimacy of centralized resource allocation, especially of players. Second, centralized managers must not suffer from asymmetric information in relation to team managers. If day-to-day experience in operating a team gives team managers information about production and demand that is not easily made available to the league manager, centralized resource allocation will be sub-optimal. Third, moral hazard problems associated with the lower-powered incentives of line employees as team operators, compared with the higher-powered incentives of team owner-managers, must not be so large that they offset the advantages of centralized resource acquisition and allocation.

Other than efficiency effects, competition law provides another motivation for single-entity leagues. Competition policy makes collusion among horizontal competitors illegal, but does not prohibit cooperation among divisions of a company. Many activities of sports leagues are collusive, such as restricting competition for players and joint marketing of broadcast rights and product licenses. Joint-venture leagues risk successful antitrust attacks if their teams collude to monopolize a product market or monopsonize an input market. One response is to obtain an antitrust exemption, either in a statute or through collective bargaining, since labor unions typically are exempt from competition laws. A single entity avoids exposure to antitrust problems without the need to seek cooperation from politicians or unions.

Historical experience indicates that the joint-venture structure is substantially more efficient than the single-entity structure. Indeed, no single-entity league has ever been successful in any sport in North America or Europe. Major League Soccer (MLS) in the U. S. is an especially interesting case. MLS did not emerge from a group of existing teams, but was created by Alan Rothenberg, who successfully organized the 1994 World Cup in the United States. Because American teams and players wanted to be eligible for international competition, a necessary condition for a successful major league in U.S. soccer was to obtain approval from the national sanctioning body, the U. S. Soccer Federation (USSF). Two groups applied to USSF for approval to start a league: a group of professional teams that were members of a joint-venture second division league, and a group of individuals led by Rothenberg who proposed a single entity. The USSF accepted Rothenberg's proposal and rejected the other group's proposal to enter as either the sole first division league or a competitor to MLS.

The original plan for MLS was a pure single entity. The board of directors was to be investors in the league, including corporate sponsors and broadcasters. The teams were to be centrally managed, with team operators being league employees. Unfortunately, MLS was unsuccessful in selling these ownership shares. Potential investors wanted to own teams, not a share of a league.

MLS then changed its structure to allow "owner-operators" of teams. These individuals buy an ownership share in the league and become members of the league's board of directors, as before, but also acquire the right to operate a team. The league retains the authority to decide precisely what authority was delegated to owner-operators, to negotiate player contracts, to assign players to teams, to sell or trade players to teams in other leagues (notably, to teams in first division European leagues), to sell national broadcasting rights and product licenses, and to negotiate stadium leases; however, the owner-operators hold all but two positions on the board, and so control these decisions. Owner-operators also are given limited rights to find talented young players who are signed by the league but allocated to their team, to hire coaches, to sell local broadcast rights and product licenses, and to handle team marketing. MLS pays the costs of players and stadiums, and in return teams give the league half of their revenues.

This structure is nothing if not innovative. The courts subsequently ruled that it passes the single-entity test, treating the independent owner-operator entities as a form of franchise rather than a group of horizontal competitors. Thus, MLS has its cherished antitrust exemption for monopsonizing the player and stadium markets.

Unfortunately, MLS is not a successful sports league. Success in monopsonizing the player market has little value because the best American players can and do obtain competitive salary offers to play in Europe. The U.S. national team now ranks as among the top dozen or so soccer powers, but most of the best players on the national team play for European professional teams. Thus, America's stars mainly play in Europe, and MLS is regarded as a minor league, comparable to the bottom of Division 1 or the top of Division 2 in England.

The centralization of most cost responsibility and revenues in the league creates low-powered incentives for team owner-operators. As a result, MLS teams have not engaged in much marketing and community relations, since most of the benefits of those actions would accrue to the league, not the team. This problem has been solved in a peculiar way: by concentrating team ownership. Two individuals, Philip Anschutz and Lamar Hunt, are the owner-operators of nine of the ten teams in the league.

Concentration of ownership exacerbates another problem – the perceived unfairness and manipulation in centralized control of teams, especially player assignments. A common owner creates the perception, regardless of the management details, that teams do not really compete. Player allocations, trades and decisions whether to retain players or to allow them to transfer to Europe are likely to be perceived as manipulations of relative team strengths for some grander corporate purpose, rather than the outcome of a competitive quest for a league championship.

MLS may be struggling because the market for soccer is weak in the U.S. But this explanation seems implausible, given the number of youth who play soccer, the interest of Americans in the World

Cup and other international competitions, and the presence in the U.S. of a very large immigrant population from nations where soccer is the national sport. A better explanation is that MLS is poorly organized, starting with its single-entity structure.

External Governance

Regardless of the details of ownership structure, sports leagues can be subjected to varying degrees of supervision and control by external organizations. An important part of the discussion of sports policy is whether external organizations ought to have a stronger governance role. The motivation behind proposals to expand external supervision is to increase the degree to which league operations reflect the interests of society in general, not just team owners.

Three possible sources of external control are possible. Governing bodies of a sport, normally non-governmental organizations, already exercise some authority at all levels of play, including both amateur and professional competitions. In addition, two other forms of external control have been proposed: a league chief executive that is independent of member teams, and a government regulatory authority to oversee sports leagues.

Most teams sports have strong non-governmental governing bodies that encompass both professional and amateur athletes, teams and leagues. Governing bodies normally assert some jurisdiction over league operations. Although sports differ substantially in the division of authority between leagues and governing bodies, most governing bodies set rules of play, rules for player eligibility and conduct (including drug use), and organizational rules for teams and leagues, thereby playing a role in league decisions about format, hierarchy, multiplicity and membership.

Professional sports leagues often modify the rules of play, eligibility and behavior of governing bodies. For example, in North America the National Basketball Association (NBA) has slightly different rules than the International Federation of Basketball Associations (FIBA) with respect to the size of the three–second area and the time on the shot clock, and negotiates its own behavioral rules for players with the National Basketball Association Players Association; however, in international play, American players and teams must abide by FIBA's rules.

The extent to which a league or other organization of teams is likely to adopt very different rules from other leagues depends on two factors. One is the extent to which teams find value in either scheduling contests across leagues or engaging in extensive exchange of players. If the rules of two organizations are radically different, cross-organization matches and player exchanges are less attractive.

The authority of governing bodies also depends on the importance of the international matches and championships that they control, including their power to determine the eligibility of teams and players in these events. Governing bodies can exercise more authority if teams and players place a high value on international competition and so would experience a substantial loss if they were disqualified from these events for disobeying the rules of the governing body. Thus, FIFA is a much stronger organization than FIBA because one professional basketball league, the NBA, dominates basketball and places relatively little value on FIBA's international events, whereas to top soccer teams and players, participation in international events is at least as valuable as participating in national leagues.

The historical record of strong governing bodies, such as FIFA and the International Olympic Committee, reveals their limitations as regulators of professional sports. The goals and incentives of these bodies differ from those of major professional leagues, but not as radically as one might suppose.

Governing bodies typically have a broader constituency, including many more organized participants in a sport, so they are likely to be more balanced in resolving disputes among major, minor and amateur leagues. But their structure gives them a motivation to suppress competition in even a broader way than might be sought by a professional league. For example, governing bodies in team sports generally favor restrictive player market rules, collective sale of broadcasting rights, and limitations on the number and size of top professional leagues (to protect lower leagues against competition).

The strongest governing bodies are those that control valuable events, such as the Olympics and the World Cup, and they are likely to be very protective of these events because their ability to carry on their broader missions depends on maximizing income from them. It is not realistic to expect that these bodies will take any action that undermines the value of these important sources of income. An example is the attack of the International Olympic Committee on other organizations that use the word "Olympics." Another example is the opposition of governing bodies in national and international football to the formation of international major professional soccer leagues, such as the requests of Wimbledon to move to Dublin and the Old Firm, Celtics and Rangers, to transfer from the Scottish Premier League to either the English Premier League or the Dutch first division. Permanent international leagues would undermine the financial value of existing competitions and hence the authority of governing organizations.

Because governing bodies are unlikely to regulate professional leagues according to a broader public interest, some have proposed either a strong, independent commissioner or a government regulatory authority. As with stronger governing bodies, the likely effects of these proposals depend on the goals and incentives of the authority. The fundamental problem with both proposals is that institutionalizing an authority that reasonably balances all interests, including those of fans and even citizens who are not fans, is extremely difficult. Regarding an independent non-governmental commissioner, the crucial issue is the selection process: who hires and, if necessary, fires the commissioner? Whereas one can imagine institutionalizing a process whereby some external group that includes players, major-league teams, minor-league teams, and amateur teams picks the chief executive of professional leagues, going much beyond this list of interested parties is likely to make the process very cumbersome, yet excluding fans, broadcasters and other customers and suppliers of professional sports creates a strong likelihood that the chief executive will take actions that disadvantage unrepresented groups. For example, the chief executive would have an incentive to establish anticompetitive policies aimed at unrepresented groups in order to benefit the represented constituencies.

Government regulation presumes that a combination of political appointees and civil servants will oversee the operation of leagues. The main problem with this approach is the inefficiencies arising from the substitution of administrative rules for market incentives. Regulation requires that society devote resources to setting and enforcing rules. Moreover, because regulatory rules can have a significant effect on the income of regulated entities, it requires cumbersome procedures to ensure the fairness of rules that are costly to obey. As a result, regulation has a tendency to slow down decision making and to make regulated entities less flexible in responding to new circumstances.

Because regulation is costly, imposing it is likely to be socially beneficial only if an inefficiency arises in a market that can not be dealt with by restructuring the market or creating appropriate financial incentives. Rule-making regulation is intended to constrain the extent to which regulated entities pursue financial incentives. Most of the problems associated with the operation of sports leagues arise from the fact that leagues are monopolies. The one source of inefficiency that is not a manifestation of market power is the externality arising from the presence of a team, but the most effective remedy for inadequate provision of a public good is a subsidy, not a regulation that demands expansion of a sport beyond the level at which sports enterprises are financially unviable. Thus, a necessary condition for regulation of sports leagues to improve welfare is that competition is not feasible. Once again, an issue of sports policy turns on whether leagues and teams are natural monopolies, so that using the market incentives arising under competition is not feasible.

NATURAL MONOPOLY

In America, major league team sports are organized into separate leagues or divisions, but almost all metropolitan areas have only one team. In Europe, major league team sports tend to have only one undivided league. The exceptions are the European Champions League and Scotland's two-tiered Premier League. But large cities tend to have several professional teams. While this pattern calls into question whether either teams in a city or major leagues in a sport are natural monopolies, owners and journalists often claim otherwise. This section examines whether teams or leagues are natural monopolies, and the implications for the membership rules of leagues that emerge from this analysis.

Teams

Leagues that practices promotion and relegation generally do not give their members exclusive territorial rights. Thus, teams in closed leagues have an important advantage over teams in leagues that practice promotion and relegation. Not only are the former assured that they will continue to play in the highest league after a string of woeful seasons, they also are assured that regardless of how badly they perform, they will face no competition from entrants into their local market that might field a stronger team and win their fans.

In the United States, sports leagues and journalists have defended territorial rights by arguing that local teams are natural monopolies. The rationale is that local competition leads to two undesirable outcomes. First, because fans prefer better teams, teams in the same market will engage in an "arms race," each trying to attract better players than the others in order to capture the interests of local fans. Second, one team will succeed – perhaps through luck – in being the best, and will attract many fans; however, the other teams, despite participating in the player arms race, will suffer at the gate because they are not the best locally. This argument amounts to the claim that sports is a type of "winner-take-all" product in local markets (Frank and Cook, 1995). The also-rans, therefore, will suffer high costs and low revenues, and be driven from the market.

Considerable evidence shows that these arguments are false. In leagues without territorial rights, markets with more than one successful team have proven to be stable. In Italy, two teams from Milan played in the 2003 semi-finals for the European Cup and also finished second and third in Italian Serie A. Although less successful, Lazio and Roma from Rome have had an enduring rivalry. In 2000 the four teams in Milan and Rome were among the top five in attendance for the league.

In England, between one-fourth and one-half of the Premier League typically is comprised of teams from London. Three of these, Arsenal, Chelsea and Tottenham Hotspur, have been more or less permanent members of the league, and seven others, Charlton, Crystal Palace, Fulham, Queen's Park Rangers, Watford, West Ham and Wimbledon, all spent time in the Premier League between 1995 and

2003. During the 1990s, fourteen London area teams played in the Football League, although one, Barnet, was demoted to the Football Conference in 2001. The six largest English metropolitan areas all contain two or more Football League clubs, and at least two in every city have been regular Premier League or Division 1 members for decades. Teams in these cities have accounted for 60 percent or more of top Premier League finishes in every decade since the 1930s (Dobson and Goddard, 2001). And, the Old Firm teams in Glasgow, Celtic and Rangers, have dominated the Scottish First Division for decades.

The fixed leagues of the United States also have had successful multi-team cities. From the formation of the American League in 1899 through 1957, New York City had three baseball teams – the Dodgers, Giants and Yankees – before the first two fled to California to become temporary monopolies in the nation's second and sixth largest markets. In both cases, the cause of the departures was not that New York could only support one team, but that New York refused to subsidize new stadiums. Baseball put a new team, the Mets, in New York five years later, added second teams in both Los Angeles and San Francisco, and maintained two teams in Chicago. In the ensuing decades, teams from Los Angeles, New York and the Bay Area have persistently won more than their share of league championships, although not the Chicago teams.

This history shows that multiple major league teams can co-exist in the same city. The underlying economic reason is that an "arms race" or "winner-take-all" theory is not an accurate conceptual model of fans and teams. In large cities, teams locate in different neighborhoods and appeal to somewhat non-overlapping groups of fans, taking advantage of intercity and inter-ethnic rivalries within a metropolitan area. Fan support is sensitive to travel distance to a match. Most fans attending a game travel less than a half-hour to the stadium. In a large metropolitan area, strategic positioning of team venues, following Hotelling's model of spatial competition, gives each a separate local area in which fans will be inclined to support the nearby team. The demand for each team will still be responsive to team quality, but not primarily because its fans will switch loyalties in response to swings in relative team quality. Thus, in large cities, teams compete for fans, but the main determinant of their success is how well they do against the league in general, not in comparison to cross-town rivals.

Table 1 illustrates this point by comparing the attendance of the Oakland Athletics and San Francisco Giants from 1995 through 2002. In 1995 and 1996, attendance in American baseball was down because of fan reaction to the 1994 labor lock-out that forced cancellation of the World Series. In these seasons, both teams played poorly, being out of the pennant races in the middle of the season, and both finished near the bottom in attendance. In 1997 and 1998, the A's continued to play poorly, but the Giants blossomed into pennant contenders. A's attendance increased slightly as fans forgave the lock-out, while the Giants attendance increased by nearly 40 percent over two years. Since 1999, both teams have been pennant contenders. In 1999, the Giants experienced another increase in attendance, while A's attendance jumped 20 percent. In 2000, the Giants moved to a new stadium, continued to be pennant contenders, and reached the World Series in 2002. The new stadium caused a huge jump in attendance, averaging more than one million fans per year more than their teams of equal quality in the old venue. Meanwhile, the A's attendance in the 2000-2002 period has averaged 50 percent higher than in 1995-1998. In 2002, both teams experienced lower attendance, despite having their best years in over a decade, due to the recession in the information technology sector, which hit the Bay Area very hard.

Table 2 shows a similar comparison, this time for Sheffield United and Sheffield Wednesday during the 1990s. The table compares both average attendance at league games and total season gross revenue, which includes matches in various cup competitions. Average attendance is used because the number of games played per season varied during the time period shown. As the table shows, attendance and revenue for these teams each move in the same direction in all but one year. When United was demoted to Division I for the 1994-5 season, its attendance and revenue fell by more than twenty percent. If ever United fans would be motivated to switch loyalty to Wednesday, the 1994-5 and 1995-6 seasons would have been the time. Yet Wednesday's attendance and revenues fell over the same period. The lost fans and revenue for United were lost from the league, not to its city rival.

As is apparent from these examples, success by one team does not necessarily harm the other. The success of a team is driven primarily by its performance, local economic conditions, and the public's general attitudes about the sport. Cross-town competition exists, but it is dominated by other factors. Hence, demand does not exhibit the "tipping" phenomenon of winner-take-all markets, and teams do not have an incentive to engage in an arms race.

Television is now a major source of income in sports. Broadcasting increases the potential for direct competition among teams; however, whether it actually increases competition depends on league policies and competition law, as discussed in the article by Tom Hoehn and David Lancefeld. Broadcasts cover at least an entire metropolitan area, and can encompass nations and even multi-national regions. If each team markets its own broadcasting rights, teams within a metropolitan area will compete for local rights, but many more teams compete for broader rights. As with attendance, fans prefer to watch their nearby local team, but distance is less likely to inhibit the creation of fan attachments to teams through broadcasts, so a strong team can attract audiences almost anywhere.

In an American closed league with territorial rights, teams in the largest cities have an attendance advantage over teams in smaller markets, and so, if well managed, are likely to be persistent winners. If these teams also sell their own TV rights and pocket the proceeds, the revenue gap between teams in big and small markets is larger. In a promotion-relegation league, teams in the same city may compete for the sale of rights, which limits the extent to which television increases the revenue gap – and hence the quality gap – between teams in large and small cities. If leagues are allowed to centralize TV rights sales, they can close the revenue gap entirely by sharing broadcast revenues equally.

The conclusion to be drawn from the preceding discussion is that teams are not natural monopolies, and so European Leagues have the more efficient policy regarding territorial rights. Market-driven decisions by individual teams about whether to enter the sport, and if so, where to locate and what level of the hierarchy to seek to attain, are a more efficient system for allocating teams among cities. In both systems, teams from larger cities tend to dominate; however, in a promotion-relegation system, league championships tend to be spread among multiple teams within large markets, whereas in closed leagues they are more likely to be concentrated in a smaller number of teams with exclusive or near-exclusive rights in those markets. The fact that American leagues tend to have only one or two teams in even the largest markets is the consequence of inefficient league rules, not natural monopoly.

Leagues

Like teams, leagues sometimes claim that they are natural monopolies. The basis for this claim is the poor survival record of competition between rival major leagues. In Europe, national sponsoring organizations have not permitted competing major leagues. In the U. S., leagues have entered frequently, but only two – the American League in baseball (1899) and the American Football League (1960) survived, and both of these merged with the incumbent.

In the past sixty years, the CFL and five other new leagues have entered American football in the U.S.: the All-American Conference (1946-9), the American Football League (1960-9), the World Football League (1974-5), the United States Football League (1983-5) and the Xtreme Football League (2001). In basketball, three leagues competed in 1946-7: the Basketball Association of America, the National Basketball League, and the American Basketball League. Over the next two years, these leagues merged into the current National Basketball Association. Later the American Basketball League (1962-4) and the American Basketball Association (1968-76) also entered. Hockey had one entrant: the World Hockey Association (1972-9). Baseball has had no entry since the Federal League (1914-5).

No entrant has survived as an independent entity. All ten AFL teams merged into the incumbent National Football League in 1970. Three AAFC teams (Baltimore, Cleveland and San Francisco) joined the NFL. The merger of the three basketball leagues left primarily teams from the BAA; nearly all of the other teams were disbanded. Later, the ABA placed four teams in the NBA and four WHA teams joined the NHL. No league that was formed since 1972 has succeeded in placing even one team in the incumbent league. Thus, the evidence is consistent with the hypothesis that a major league is a natural monopoly.

At the top level of the sports hierarchy is a natural monopoly in that only one annual national champion can be crowned. To stake a valid claim, the championship must be open to all major league teams. Sports fans express demand for championships, and this demand can be satisfied only if some institution has the authority to organize such an event. The championship organizer must have another all-encompassing authority: the right to set the rules of the contest and to schedule the events that will crown the champion.

These requirements do not support the conclusion that the top league in a sports hierarchy must be a monopoly. International championships are not organized by a single league, but use the results of numerous leagues to determine qualifications for the championship. Likewise, intercollegiate sports in the United States feature national championships, but the six leagues whose champions automatically qualify are independently operated. Thus, once leagues agree to the same playing and eligibility rules and to collaborate on a "champion of champions" event, no further business cooperation is required.

The value of separate leagues is born out by the decisions by all major American sports to organize the monopoly major league organization into separate playing divisions. As explained above, several different mini-leagues enhance fan interest by creating more championships and, if schedules are unbalanced, more uncertainty about the actual relative strengths of the teams. A single balanced league makes abundantly clear early on which few teams have any serious chance to win the championship, whereas a series of divisions with unbalanced schedules stretches out the portion of the season in which many teams appear to be in contention. In sports less information is a good thing in that it increases perceived uncertainty of outcomes, and hence demand. The American NFL has mastered the art of manufacturing uncertainty by adopting a schedule in which weak teams from the previous season play each other in the unbalanced part of the schedule, thereby bloating the record of the best among them, while the strong teams are matched against each other to bring their records closer to the rest of the pack.

If multiplicity in leagues is financially attractive, why do leagues merge to monopoly? The answer is that monopoly leagues enjoy market power in both input and output markets. In the U.S. and U.K., monopoly leagues take advantage of permissive competition policy to bargain as a unit for national television contracts. A monopoly league also may be allowed to enforce restrictions on competition for players among teams. Monopoly leagues also can create scarcity in league memberships, which drives up the value of franchises.

In the United States, for fifty years teams have made use of relocation as a way to escape cities that would not subsidize a stadium. A necessary condition for relocation to be feasible is that a major league must leave some attractive markets without a major league team. Thus, to enable teams to increase their value by moving or threatening to move, a league must create an opportunity for a new league to enter.

A few years before the Dodgers and Giants moved to California, the AAA Pacific Coast League, which fielded strong teams and was very successful at the gate, declared its classification as "open" and threatened to claim major league status. Had baseball practiced promotion and relegation, the Los Angeles Angels and San Francisco Seals would have been in the major leagues decades earlier, the relocation of the Dodgers and Giants would never have occurred, and the PCL would never have been a credible threat to become a third major league. By relocating the Giants and the Dodgers, baseball removed the threat of a third major league by carving out the heart of the PCL.

A promotion and relegation system also discourages the formation of competitive leagues without using team relocation. Europeans argue that "cultural differences" explain why European football squads do not move, but American teams do. But the culture shock of losing a team is every bit as harsh in the United States. In 2003, still another best-selling book was written on the 1957 departure of the Dodgers from Brooklyn, along with an argument about who was to blame.[12] Thus, the economically most plausible interpretation of the differences between Europe and the U.S. is that the promotion and relegation system provides a less disruptive way for leagues to reorganize their membership in response to geographic shifts in population and wealth than the relocation mechanism that is substantially more common in leagues of fixed membership.

In both types of leagues, the primary danger is that the top league will restrict membership so severely that a new top league is commercially viable. Like the PCL in the 1950s, the English Division I could withdraw from the promotion and relegation arrangement with the Premier League and declare itself to have equal status – except that the governing bodies of the sport would never permit it. Moreover, its strongest members probably would oppose the change because they anticipate earning Premier status in the near future.

TEAM INCENTIVES UNDER PROMOTION AND RELEGATION

In addition to effects on entry, league size and team locations, the promotion and relegation system has other important effects that arise from the incentives it creates for teams.

The essay by Stefan Kesenne discusses competitive balance: the extent to which teams are sufficiently closely matched that game outcomes are uncertain. Promotion and relegation affects the distribution of team quality (Ross and Szymanski, 2001, and Noll, 2002). If a team receives a net financial benefit from promotion, and experiences a financial loss from relegation, teams at the bottom of a league have a greater incentive to improve team quality than would be the case in a league of fixed membership. The reason is that an improvement in team quality today not only increases revenues in the current season, but increases the probability of enjoying greater profits in the next season. For teams at the top of the best league, there is no opportunity for promotion, so there incentives are unaffected by the fact that some weak teams will be relegated. Thus, the net effect is to reduce the quality gap between the best and worst teams.

For lower leagues, the effect on relative balance is theoretically ambiguous. Unlike the top league, the best teams in a lower league can be promoted, which provides an extra incentive to improve their quality because of its effect on the probability of future profits in a higher league. Thus, qualitatively the same effect operates on teams at the top and bottom of a lower league. Whether balance is improved depends on the relative size of the profit incentives to obtain promotion versus to avoid demotion. Plausibly, the incentive to achieve the Premiership is greater than the incentive to avoid relegation to Division 2, and if so, promotion and relegation increase the spread in team quality in Division 1.

Promotion and relegation also affect demand. For teams at the top of a lesser league, the prospect of gaining promotion may enhance interest in end-of-season games. While this effect may not be very great for the teams at the very top of the standings that are fighting for the championship, for teams that are near the top but out of contention the prospect for promotion increases the stakes of late-season games. In addition, the modern format that calls for a playoff for the last promotion spot provides another opportunity for increased revenue.

End-of-season games between teams at the bottom of the league standings do not affect the championship race, which ordinarily would reduce demand. But if these teams are fighting to avoid relegation, these games become consequential, and teams seeking to avoid demotion have an incentive to put forth best efforts in trying to win. Both effects are likely to enhance demand.

Promotion apparently has a beneficial long-run effect on the demand for teams that are subsequently relegated. Studies of the demand for sport find that attendance in the previous season has a fairly strong carry-over effect into the next season. Although the focus of most of this research has been the carry-over effect of championships, the same effect apparently is present after promotion and relegation. Demand studies show that the battle for obtaining promotion increases a team's attendance, and then the act of being promoted increases it still more. Interestingly, although subsequent relegation depresses attendance, the magnitude of this effect is smaller than the magnitude of the effect of promotion. Consequently, in Year 3, a team that was promoted after Year 1 and relegated after Year 2 will have higher revenues than a team that had exactly the same record in Year 3, but was never promoted or relegated (for examples, see Noll, 2002). The Premier League adds to this benefit by providing "parachute" revenues to teams for two years after they are relegated.

The long-term revenue effect from promotion explains what in England have been called "yo-yo teams" – squads that regularly bounce back and forth between two leagues. In some cases, bouncing between leagues apparently is planned, in that some teams spend more on players in the year that they are fighting for promotion than in the next year in the stronger league.

The significance of these demand effects is that instituting promotion and relegation increases fan interest and hence team revenues, compared to the interest and revenues that would be expected in a league of fixed membership. Thus, leagues that practice promotion and relegation generate more economic welfare than fixed membership leagues.

Nevertheless, the promotion and relegation system probably reduces profits, especially among the best teams in the top league. Promotion and relegation increases the value of the best players to teams that are in battles for either promotion or relegation. If the supply of high-quality players is not perfectly elastic, this increase in demand will cause player wages to be higher. Whereas teams in battles for promotion or relegation receive some additional revenues, the teams at the top of the best league do not -- promotion and relegation does not affect there demand, so they end up paying more for talent without a compensating benefit.

CONCLUSIONS

The organizational structure of leagues is a fundamentally economic decision that affects the extent of competition in sports, the distribution of teams in a nation, the relative playing strengths of teams, and, most importantly, the amount of economic benefits that a sport delivers to its customers. The analysis in this article leads to three main conclusions.

First, promotion and relegation systems are superior to fixed membership leagues as a means of delivering economic benefits to fans. Promotion and relegation improves competitive balance at the top, and has several other demand-enhancing effects. Moreover, it facilitates distributing teams across locations in a manner that delivers greater consumer benefits than does a system of fixed memberships.

Second, competition is desirable in sports, despite the pleas of sports enterprises and the pliancy of governments in granting sports special status in competition policy. Neither teams not leagues are natural monopolies: stable competition among teams in a city and among leagues in a nation is feasible and desirable, and serves to equalize revenues among teams in the absence of entry restrictions. Government acquiescence in the decisions by governing bodies to form monopoly leagues and antitrust exemptions for these leagues are policy mistakes.

Third, all existing organizational forms produce two few teams at the top level. All nations allow the top leagues to be monopolies, and monopolies leagues contrive scarcity in teams to increase the wealth of incumbents. Leagues achieve this objective by effectively erecting barriers to entry against new leagues and by restricting their own size for the purpose of maximizing the profitability of existing members. The incentive to restrict membership is especially strong if leagues engage in significant revenue sharing.

Fourth, multiple pathways to championships and nonexclusive league membership deliver benefits to consumers by increasing the number of successful teams and the importance of each match. Here both Europe and North America have something useful to learn from each other.

For Europe, the lesson is to fragment leagues and introduce post-season playoffs, expanding on the current practice of using playoffs to determine the last spot for promotion. One can imagine a four-division Premiership with five (or more) teams in each division and an unbalanced schedule among divisions, followed by a championship tournament of the divisional winners. This format would reduce the number of League contests by several games, freeing space for the post-season tournament and perhaps for other match formats.

For the United States, the lessons are two. One is to adopt the promotion and relegation system, freeing minor league teams from perpetual secondary status. The other is to introduce season-long national tournaments, like the FA Cup and FL Cup, as substitutes for some regular season games in the excessively long seasons in baseball, basketball and hockey. Certainly sacrificing eight games to return to a 154 game schedule in baseball to make room for a summer tournament would enhance revenues, especially from broadcasting.

Perhaps the greatest lesson to be learned from the study of comparative leagues is the amazing variety across sports and countries in how basically the same types of businesses have attempted to solve the same basic puzzles. Comparative study of sports leagues is a useful exercise. It tends to explode cherished myths born from narrow experience with one's favorite sport, and it provides substantial insights into the incentives that lie behind both the good and bad features of professional team sports in every nation.

Table 1: Attendance and Winning Records of Oakland and San Francisco

	Oakland A's		***San Francisco Giants***	
Year	***Fraction Won***	***Attendance***	***Fraction Won***	***Attendance***
1995	.465	1.17	.465	1.24
1996	.481	1.15	.420	1.41
1997	.401	1.26	.556	1.69
1998	.457	1.23	.546	1.93
1999	.537	1.43	.531	2.08
2000	.565	1.73	.579	3.34
2001	.650	2.13	.556	3.31
2002	.636	1.89	.590	2.89

Note: attendance is in millions.

Table 2: Attendance, Revenues and Winning Records in Sheffield

		United			**Wednesday**	
Year	***Finish***	***Attendance***	***Revenue***	***Finish***	***Attendance***	***Revenue***
1992	P-9	21,805	Ƚ 4,265	P-3	29,578	Ƚ 7,516
1993	P-14	18,801	6,060	P-7	27,263	12,806
1994	P-20	19,562	5,431	P-7	27,191	11,914
1995	I-8	14,462	4,325	P-13	26,572	10,995
1996	I-9	12,901	4,311	P-15	24,877	10,078
1997	I-5	16,638	5,133	P-7	25,693	14,335
1998	I-6	17,942	8,536	P-16	28,709	16,303
1999	I-8	16,243	6,421	P-12	26,745	19,124

Note: position indicates league (Premier - P, Division 1 - I) and finishing place in standings, attendance is per league game, and revenue is thousands of pounds.

REFERENCES

Cairns, J. A. (1990). "The Demand for Professional Team Sports," *British Review of Economic Issues* 12, pp. 1-20.

Dobson, Stephen, and John Goddard (2001). *The Economics of Football*. Cambridge: Cambridge University Press.

Frank, Robert H., and Philip J. Cook (1995). *The Winner-Take-All Society*. New York: The Free Press.

Hoehn, Thomas, and Stefan Szymanski (1999). "The Americanization of European Football." *Economic Policy* (April), pp. 205-40.

Noll, Roger G. (2002). "The Economics of Promotion and Relegation in Sports Leagues: The Case of English Football," *Journal of Sports Economics*

_______ (forthcoming). "The Economics of Contraction in Baseball," *Journal of Sports Economics*.

Noll, Roger G., and Andrew Zimbalist (1997). *Sports, Jobs, and Taxes*. Washington, D.C.: Brookings Institution.

Shapiro, Michael (2003). *The Last Good Season.* New York: Doubleday.

Szymanski, Stefan (forthcoming). "The Economic Design of Sporting Contests." *Journal of Economic Literature*.

Zimbalist, Andrew (2003). *May the Best Team Win*. Washington, D.C.: Brookings Institution.

FOOTNOTES

* Morris M. Doyle Professor of Public Policy, Department of Economics, and Director, Stanford Center for International Development, Stanford University. The author grateful acknowledges the useful comments on an earlier draft from Ken Mayhew, Stefan Szymanski, Andrew Zimbalist, and two anonymous referees.

1. The two alternatives to league organization are "independents" and "barnstormers." An independent is a team that does not belong to a league but creates an annual schedule of matches with other teams (some of which may be league members). Most team sports began in this way, but few true independents now exist anywhere. A notable independent is Notre Dame Univeristy in American intercollegiate football, which usually is among the best college teams and always ranks at or near the top in revenues because of its very lucrative long-term television contract. Barnstormer is the name given to teams that have no regular home field, and that travel the country playing exhibitions. The opponent is sometimes another barnstormer that travels with them, with the leading example being the Harlem Globetrotters basketball team and their perennial patsies, the Washington Generals. For other barnstormers, the opponent is a team of local all-stars from the community in which the game is played. In both cases, the games are usually mis-matches, with the barnstormers vastly superior to their opponents. The fundamental reason that nearly all teams play in leagues is economic: the demand for games that are part of a championship schedule vastly exceeds the demand for exhibitions, or games that determine nothing more than the identity of the winner of a particular game.

2. For surveys of research on the demand for team sports, see Cairns (1990) and Dobson and

Goddard (2001), Chapter 7.

3. Although the Premier League is now separate from the Football League, to avoid repetitious explanation this article will refer to all four top English leagues as the Football League. In addition, the new designations – Premier League and Divisions 1, 2 and 3 will be used to identify the four leagues in the period when they were known as Divisions 1, 2, 3 and 4, respectively.

4. For much of the 20th century, baseball had three further classifications -- B, C and D, but no Rookie classification; however, these were eliminated in the 1960s after television eroded the demand for minor league baseball. The new Rookie classification applies to leagues with a much shorter season in which players who have signed their first professional contract are given instruction prior to playing their first full professional season, usually at the A level.

5. Obviously, such a mega-league could not schedule every team to play every other, and so would be organized as several divisions. The analysis here assumes that each division would have roughly equal representation from each of the four existing leagues – five Premier members and six each from the other three divisions. In this arrangement, a team that currently plays in Division 3 would have two or three home games per season against a team that is in contention for the division and league championships, and all the other games, especially once the season is well enough along for the contending teams to become clear, would have no significance in terms of the championship race. In many of these games, the team would be grossly over-matched, even though the opponent is not a championship contender.

6. The five leagues are: the Continental Basketball Association (which has operated since 1947 except for 2001), the National Basketball Development League (which is owned by the National Basketball Association), the United States Basketball League, the United Professional Basketball League and the Eastern Basketball Alliance. Two others, the American Basketball Association and the Xtreme Basketball League, folded during the 2002-3 season.

7. The AAA leagues were the American Hockey League and the International Hockey League (IHL). The AA leagues were the Central (CHL), East Coast (ECHL), United, West Coast (WCHL) and Western Professional Hockey (WPHL) Leagues.

8. The IHL folded, the WPHL merged with the CHL, and the WCHL merged with the ECHL.

9. Whether the system actually determines a clear national champion is sometimes a matter of controversy. Some undefeated teams from lesser conferences have been excluded from the BCS, and in when two teams tie for a league championship the team that proceeds to the BCS has been decided by a coin flip or an arbitrary rule, such as which team went to a major bowl most recently.

10. For an analysis of why the top teams nevertheless may withdraw from national leagues and form a European Superleague, see Hoehn and Szymanski (1999).

11. The NFL is unique: of the 32 teams presently competing, thirteen came from mergers. In the past fifty years, the NFL has added only seven teams by expansion.

12. Shapiro (2003). The villain in this book is Robert Moses, who controlled New York City's

redevelopment funds and refused to subsidize sports facilities as part of urban renewal.