

SPORTS ECONOMICS AT FIFTY

by

Roger G. Noll

Department of Economics
and
Stanford Institute for Economic Policy Research
Stanford University

In 2006, sports economics celebrated its fiftieth birthday. This essay reviews Simon Rottenberg's seminal 1956 article and a handful of articles appearing over the next fifteen years that established the conceptual foundation for the field. A common thread among these papers is that all were motivated by the same policy issue: whether professional sports leagues need exemption from antitrust laws (especially with regard to player markets) in order to operate efficiently. These papers undermined the case for antitrust exemptions, and were the sources of the arguments that brought free agency and competition to the markets for professional athletes around the world. Moreover, although these papers lacked serious empirical analysis and, with one exception, did not include formal theoretical models, in the subsequent thirty-five years most (but not all) of the original empirical and theoretical arguments still stand.

Forthcoming in Plácido Rodríguez, Stefan Kesénne and Jaume García (eds.), *50 Years of Sports Economics*.

SPORTS ECONOMICS AT FIFTY

by

Roger G. Noll

The economics of sports was born in 1956 with the publication of Simon Rottenberg's classic paper. After only a handful of subsequent publications in scholarly economics journals during the next 15 years (El-Hodiri and Quirk, 1971; Jones, 1969; Neale, 1964; and Sloane, 1969 and 1971), the field blossomed in the 1970s and has grown steadily since. Fifty years after the publication of Rothenberg's article, sports economics has a professional society (the International Association of Sports Economists) and two dedicated scholarly journals, *The Journal of Sports Economics* and *The International Journal of Sports Finance*. Reportedly over 100 U. S. colleges offer courses dedicated exclusively to the economics of sports (Kirchhoff 2006), and world wide at least ten textbooks are devoted exclusively to sports economics (Andreff and Nys, 1986, 2002; Bourg and Gouget, 1998; Buch and Schellhaas; Downward and Dawson, 2001; Fort, 2003; Heinemann, 1995; Leeds and von Allmen, 2005; Li, Hofacre and Mahoney, 2001; Sandy, Sloan and Rosentraub, 2004; and Trosien, 2003).¹

The purpose of this essay is to evaluate the intellectual history of the research literature on the economics of sports, starting with Rottenberg's seminal article, and along the way to offer some new analysis and conjectures about where the field might be going in the future. Of course, because of its focus on the early years of the field, this review is incomplete and of

1. If one includes books devoted to the broader field of sports and recreation, several other textbooks are available.

necessity gives inadequate attention to some areas, notably the growing literature on the economics of individual sports.

Sports as a Field in Economics

The existence of sports economics as a field is problematic because it runs counter to how contemporary economists define fields. Sports is an industry, and economists generally have not regarded industries as fields in economics since traditional institutional economics was supplanted by modern neoclassical microeconomics in the 1930s.² Among economics departments that offer courses in sports, one finds no courses in automobiles, oil, steel, or nearly any other industry. Specialization among economists rarely goes beyond sectors, such as agriculture, energy, finance or transportation, and except for finance even this broader specialization accounts for a dwindling minority of economists.

The emergence of the field of sports economics can not be explained by the economic importance of the industry. The industry classification “performing arts, spectator sports, museums and related activities” (SIC code 711) produces less than one-half of one percent of U. S. value added (U. S. Bureau of Economic Analysis, 2006). More refined value added data are not available, but other data are available for the narrower category “sports teams and clubs” (SIC code 711211), which is the focus of the vast majority of publications in sports economics. Team sports account for 23 percent of total sales in SIC 711.³ In 2002 U. S. team sports took in

-
2. The *new* institutional economics is an interdisciplinary field that studies the incentive structures of institutions (see <http://www.isnie.org/>); however, like the rest of modern economics, this field develops general methods that economists regard as widely applicable.
 3. U. S. Census Bureau (2004b), Tables 2 and 3.

less revenue than golf courses (which are measured separately from golf tournaments), gambling casinos, fitness and recreation centers, canned fruits and vegetables, or animal food. In the same year employment by sports teams was about 40,000, compared with 122,000 for theatrical companies, 441,000 for fitness centers, 82,000 for bowling alleys, 311,000 for golf courses, 49,000 for musical groups and artists, 92,000 for museums, 111,000 for amusement and theme parks, 47,000 for animal food, and 59,000 for fruit and vegetable canning.⁴ Needless to say, one searches in vain for college courses on the economics of fitness centers, pet food, or canned fruits and vegetables.

Despite its meager conventional economic significance, the reasons for the development of sports economics are transparent. Foremost is that sport has far greater social significance than can be measured by employment and sales. If the criterion for evaluating the industry is the attention accorded to it by mass media such as newspapers, general interest magazines, and broadcasters, sports ranks among the top two or three industries. The widespread popular interest in sports extends to academic economists. Notwithstanding outward appearances, many economists are human beings with conventional interests. Economists are among the lucky few who can combine business with pleasure by making sports an element of a career.

A second reason for studying sports is that it offers interesting theoretical and empirical challenges in applied microeconomics. The importance of the major scholarly contributions to the field arises from peculiarities in the sports production process and the demand for sporting events. The importance of this work to society is due to the array of public policy issues that

4. These data are compiled from the following publications of the U. S. Bureau of the Census (2004a, 2004b, 2004c, 2004d, 2004e).

arise in the industry, including competition policy, broadcast regulation, labor relations, and government subsidies. The focus of this essay is on this second reason for the emergence of the field in the last few decades.

Origins: 1956-71

Six early publications are widely known among sports economists because they were the first to introduce important issues that motivated substantial subsequent research. These six publications are not the only writings by economists on the topic; however, they are the only publications that are devoted exclusively to sports that appear in leading scholarly journals. Several other publications mention sports to illustrate a more general point, but do not treat an issue in sports economics in depth and so have not been as influential in developing the field. Examples are Bain (1967), Becker (1964), Gitelman (1968) and Weintraub (1958). Also omitted are publications by economists that are not primarily economics, such as Andreano (1965) and Ritter (1966), and works in other disciplines, most notably history and sociology, that address economic issues. Finally, I omit government reports that were written all or in part by economists on the grounds that these are not subjected to the same professional vetting as scholarly books and articles. Nevertheless, government reports played an important role in the field by providing both motivation and empirical information for subsequent scholarly work, as demonstrated by the citations in the early publications.

The choice of an ending date for the early period inevitably is arbitrary, but I have an excuse. In December 1971, the Brookings Institution sponsored the first conference on the economics of sports in which a dozen research papers were presented to an audience that included not only most of the economists who had done research on sports but also owners of

sports teams, league executives, heads of players associations, and professional players who had graduate training in economics. This conference more than doubled the number of research papers in the field, launched several new threads of research, and led to a book that includes several papers that still are cited frequently in the literature (Noll 1974). I have made this conference the boundary for the origins period because it is the natural founding date when sports economics made the transition from the topic of an occasional paper to a field of research.

The early works in sports economics were motivated by a practical concern: whether the financial viability of teams requires special institutions and public policies, especially with respect to the labor market for professional athletes. When economists began to study sports, all top level (“major” and “first division”) professional sports teams in the world were members of a monopoly organization that operated under rules severely restricting competition among teams for players. Owners justified these rules by arguing that they were needed to create “competitive balance” among teams, which in turn was necessary to make customers (fans) interested in the sport. The argument was that in the absence of player market restrictions, the best teams would acquire such dominance that the lesser teams would have no chance of winning, which in turn would cause all teams to lose fan support and, therefore, leagues to collapse. For the most part, governments at the time accepted this argument, notwithstanding that in the industrialized societies of the West league practices conflicted with competition and labor laws.

Probably the most important contribution of sports economics has been its analysis of the relationship between player market rules and the performance of teams and leagues, both on the field and in the annual financial report. Although economists have not reached unanimous agreement on any key issues in sports economics, the consensus view is that the argument put

forth by owners to defend player market restrictions was incorrect and self-serving. This consensus played an important role in shifting public policy in favor of liberalized player markets. The main arguments supporting this policy shift found their first convincing expression in the handful of scholarly publications that were published between 1956 and 1971.

Rottenberg

Rottenberg's contribution to the economics of sports, contained in a single article, is that he provides the first clear explanation of how the market for players works and why monopsonistic restrictions in that market do not cause on-field competition to be more balanced.⁵ In making his argument, Rottenberg accepted the core premises of the owners' argument for the necessity of monopsony in the market for players, and argued that their conclusions did not flow from these assumptions.

Regarding the demand for sports, the core assumptions are that the demand for sporting contests is greater in a particular locality if a contestant is associated with that locality (a "home team"), increases with the quality of the home team and the uncertainty of the outcome, decreases as admissions prices increase, and depends on local conditions such as income, population, characteristics of the stadium, and the number of locally available substitutes. Rottenberg summarized these hypotheses as implying that team revenue functions have two important features. First, for each team revenue is maximized at a won-loss record above .500, but that at some point improvements in team quality experience diminishing returns and perhaps even turn negative. Second, due to other factors that affect demand, markets differ in the

5. For excellent 50th anniversary explanations, evaluations and celebrations of Rottenberg's work, see Fort (2005), Sanderson and Siegfried (2006) and Sloane (2006).

revenue that a team of given quality can generate, implying differences among teams in the incremental revenue that arises from an increment to team quality.

Regarding the demand for players, Rottenberg assumed that teams maximize profits, and that as a result a team's willingness to pay for a player is derived from the demand for contests. Profit-maximizing teams acquire players up to the point at which the marginal cost of talent equals its marginal revenue. Combining this analysis with the assumptions about market demand leads to the conclusion that if all teams are of equal quality (and expect a .500 won-loss record), the willingness to pay for a star player is higher for a monopoly team in a better market than for a team in a market in which demand is less intense.

Regarding players, Rottenberg assumed that the supply of players is elastic, but that some players have superior skills and do not have perfect substitutes. He also assumes that all players have a reservation wage – reflecting their best employment opportunity outside of sports – that is below their marginal revenue product in sports, due to roster limits and other restrictions on the number of players who are employed. Finally, he assumes that teams have monopsony rights in players, but buy and sell player contracts in a competitive market.

From these assumptions, Rottenberg derives what others subsequently called the *invariance hypothesis*: the allocation of players among teams under a monopsonized player market is the same as under competition. The logic of this conclusion is as follows. Suppose Team A values a player more highly than the player's current employer, Team B. In a competitive players market, salary is the player's marginal revenue product with Team B, but Team A is willing to offer the player a higher salary because the player's marginal revenue product is higher with A. In a monopsonized labor market, Team A is willing to buy the player

from Team B at any price below the difference between the player's marginal revenue product for A and B. In either case, a mutually beneficial bargain can be made, so the player changes teams. Hence, concludes Rottenberg, the only effect of monopsonizing the player market is to transfer wealth from players to teams, not to improve competitive balance among teams.

Rottenberg extends his analysis to include the possibility that acquiring and maintaining playing skills requires costly training. He explains that the incentive to engage in training is the same regardless of player market rules. Under monopsony, teams have an incentive to train players, while under competition players have the same incentive to pay for their own training. This observation leads Rottenberg to conclude that the amount of training is unaffected by the allocation of the property right to the player's services. As explained below in the discussion of the relationship between Rottenberg's argument and the Coase Theorem, Rottenberg's analysis is not correct without further assumptions, and under plausible assumptions training will differ under monopsony and competition.

Rottenberg notes that equal sharing of revenues among teams equalizes talent, but in so doing eliminates the incentive to seek quality, thereby undermining the demand for the sport. He also explains why a binding salary cap does not equalize playing strength, in part because teams still can buy and sell player contracts and in part because players, all else equal, will prefer to sign with better teams if salary offers are the same. Rottenberg goes on to argue that leagues can increase competitive balance by making the demand for contests more equal among teams by allowing more teams to locate in the best markets.

Rottenberg also inquires whether under monopsony players always are paid their reservation wage, as opposed to capturing a share of their superior productivity. Rottenberg

observes that the best players are paid substantially more than plausible reservation wages, with the excess correlated with performance, not other employment opportunities outside sports. He argues that the reason for this phenomenon is that the labor market for better players is more like a bilateral monopoly than monopsony. A team can not freely replace its best players *at the cost of the suppressed salary* because it must buy the rights to a substitute from another team. If a player is better than the team's second best player at the same position, the better player has bargaining power in salary negotiations, and the bargaining outcome is likely to be that better players are paid more than their reservation wage but less than their marginal revenue product.

Finally, Rottenberg addresses a proposition that he attributes to Gary Becker, which Becker (1964) eventually published with greater elaboration: whether the profits derived from salary suppression of major league players would be competed away by signing too many players, just as the profits of successful oil wells are competed away by drilling many dry holes in search of new wells. Salary suppression creates an incentive for teams to sign young players who might eventually become valuable major leaguers, at which point the player will generate profits equal to the amount of salary suppression. If the market for new prospects (in the jargon of sports, rookie free agents) is competitive, as it was in baseball when Rottenberg wrote his article because baseball had not yet adopted a rookie draft, each team will offer an initial wage ("signing bonus") equal to the present value of expected profits: the probability of becoming a major league player times salary suppression in the major leagues, minus training cost. Thus, while teams would make profits on those players whose major-league salaries are suppressed, they would lose a corresponding amount on players who were signed but failed to acquire the skill necessary to play in the major league.

Rottenberg agrees with the logic of this proposition, but argues that its effect is limited by the absence of free entry into baseball. Rottenberg's argument is not correct without further assumptions. The absence of entry prevents new teams from responding to salary suppression by entering as competitors in both the product and player markets. Consequently, the number of positions for players can not expand if more are trained to the skill level of major league players, which implies that the demand for players is inelastic; however, this argument is not really relevant to the proposition. While the possibility of growth in the demand for players due to entry of teams will effect the expected profit to be derived from signing a prospect, it will not affect the underlying equilibrium condition: new players are employed to the point at which the cost of signing and training them exactly equals the expected profits to be derived from future salary suppression, regardless of whether the player is retained or sold, and regardless of whether the market demand curve is elastic or inelastic. On this technical point, Becker is correct.

Nevertheless, Rottenberg's conclusion about dissipating profits on young players is correct for many leagues, even if his analysis is incomplete. Most leagues have other rules, many of which Rottenberg discusses, such as the rookie draft, roster limits and the waiver system, that prevent teams from dissipating the excess profits from salary suppression.

A rookie draft gives each team bargaining power over a group of players for which it has exclusive bargaining rights, with the amount of bargaining power determined by the duration of the exclusive bargaining period. This system enables teams to sign rookies at less than the competitive equilibrium wage unless the competitive wage does not exceed the reservation wage. If the probability of success were the same for all players, the supply of such players to each team were perfectly elastic, and each team could draft an unlimited number of players, then

Becker's proposition would be correct. The suppression of the signing bonus would induce every team to draft and to sign more players than it would sign if the market for rookies were competitive, which in turn would produce a market equilibrium in which each player's probability of making the major league would be lower than under competition, and the expected profit of monopsony would be zero. But if leagues limit the number of players who can be drafted or otherwise constrain the total number of players under contract, expected profits will not be driven to zero. Likewise, if a few players have a higher probability of success than the others (i.e., the supply of the most talented players is inelastic), signing a large number of lesser players will not drive the expected profits of the best players – and hence of all players – to zero.

U. S. baseball also has “waiver rules” that require teams to make players available to other teams at a stipulated price if the player does not follow a specified promotion path through the hierarchy of leagues after signing the first contract. A team that signs more rookies than it can promote to the next level within the specified period of time will lose some of them to the waiver draft. If the waiver price is below the expected value of future salary suppression, the waiver system reduces the expected profits of signing a rookie, and hence the number of players that each team will sign. In equilibrium, the waiver system then increases the probability of success of each rookie as well as increases the expected profits of teams.

As discussed and documented by Fort (2005), many economists note that Rottenberg's invariance hypothesis is closely related to the Coase (1960) Theorem. Economists sometimes mis-characterize the Coase Theorem as stating a general invariance hypothesis: in a market system, assets will find their most valuable use regardless of their initial allocation. Economists who think of the Coase Theorem in this way regard Rottenberg's contribution as a precursor to

(or special case of) the Coase Theorem because the latter is more general, or argue that Rottenberg is the inventor of the Coase Theorem and should be given at least partial credit for it.

The original Coase Theorem is far richer than a general invariance hypothesis. Coase (1960) discusses the invariance hypothesis in the first few pages of his classic paper in the context of liability rules, arguing that a choice between strict liability (the person causing harm pays) and *caveat emptor* (the harmed party bears the cost) is irrelevant if the two parties can easily contract around the problem. Coase uses the example of a cow who damages a neighbors crop by an amount that exceeds the cost of a fence, and argues that in most cases the assignment of liability determines who pays for the fence, but not whether the fence is built.

In the rest of the paper, Coase sets forth some conditions (under the rubric of transactions costs) under which the hypothesis will *not* be true. Coase argues that liability laws are a form of assignment of property rights, and that the allocation of property rights is irrelevant in some cases, but not in others, depending on the ability of parties to use market transactions or mergers to solve the problem. His main example is railroad engines that emit sparks that cause fires on adjacent land.

In the case of cows and crops, the allocation of rights – whether cows have grazing rights or crops have growing rights – does not matter, because the owner of the cows and the owner of the corn field can negotiate the least cost solution (fence out the cows or let them roam), regardless of the initial allocation of the right. In the case of railroads and burned crops, the initial allocation of the property right (railroads have the right to emit sparks or adjacent farms have the right to be free of sparks) can determine the extent of spark emissions and the extent to which crops are burned. The problem is that spark arresters are a public good for all farms

adjacent to the railroad, which can cause the transactions costs of reaching a deal about spark arresters to be so high that no transaction is possible. If initially railroads are granted the right to spark, this allocation of property rights will be final, because even if the cost of spark arresters is below the value of the crops that are burned by spark emissions, farmers are unlikely to pool their interests to pay a railroad to install spark arresters. Coase argues that the choice of institutions for solving the problem – liability laws, regulation, merger of the harmed and harming parties – depends on the circumstances.

Fort (2005) notes that Rottenberg's argument is not a “precursor” to Coase because transactions costs, which are the main focus of Coase's article, are not discussed by Rottenberg. Nevertheless, one form of transactions costs that was inspired, but not developed, by Coase pertains to agency costs. Coase (1937), which is referenced in Coase (1960), argues that firms exist because in some cases direct supervision within an organization is a less costly way to achieve efficiency than market transactions. Rottenberg's analysis of the incentive to undertake training is part of the broader story of agency theory that did not emerge in economics until long after the Rottenberg and Coase papers were published, but for which Coase is a much clearer antecedent than Rottenberg. We now know that Rottenberg's corollary of the invariance hypothesis with respect to training costs is not generally correct because his analysis failed to take into account transactions costs arising from agency theory.

Rottenberg's argument is correct under the extremes of pure monopsony or perfect competition. If all players are paid their reservation wages, a team captures all of the benefits of training that causes the marginal revenue products of players to increase. In this case, a team increases training to the point at which the marginal cost of training equals the marginal benefit,

which is the increment to a player's marginal revenue product from an incremental amount of training. But this result does not hold in actual player labor markets.

As Rottenberg notes, above some minimal skill level players are paid more than their reservation wages, which Rottenberg attributes to bargaining strength. Thus, assume that a team pays a player the reservation wage, w , or a share, s , of marginal revenue product, R , whichever is greater. Assume also that marginal revenue product is determined by training plus other exogenous unobservables, and that training exhibits diminishing returns to scale. The total expenditure on training, t , received by a player is the sum of training paid for by the team, t^* , and training paid for by the player, t^{**} . Assume that the player seeks to maximize net income, which is $w - t^{**}$ when $sR(t) < w$ or $sR(t) - t^{**}$ otherwise, while the team seeks to maximize $R(t) - w - t^*$ in the first case or $(1-s)R(t) - t^*$ in the second.

The optimal amount of total training is the value of t that maximizes $R(t) - t$, which occurs when $R' = 1$. Players who are always paid their reservation wages derive no benefit from training, so their optimal t^{**} is zero. In this case the team derives all of the benefit ($s = 0$), so the optimal value of t^* will be the optimal amount of total training.

For players who are paid more than their reservation wages, training is a public good that benefits both the player and the team, and suffers from the standard public goods problem. The non-cooperative equilibrium for players is to train to the point at which $sR' = 1$, while for teams the equilibrium is $(1-s)R' = 1$. The equilibrium amount of total training in this case will be less than optimal. For example, suppose that $s > .5$, and that players spend their individual optimal amounts on training so that $sR' = 1$. Because $s < 1$, the optimal t^{**} for the player is less than the amount that maximizes net income from the player's skills. Because $(1-s) < s$, the optimal

amount of training for the player already exceeds the optimal amount for the team, so the team will spend nothing more on training.⁶ Thus, the player market system that is described by Rottenberg produces too little training.

The preceding analysis can be reversed under two other assumptions: (1) the best training for a player depends on a multidimensional array of skills, and (2) teams, because they observe and employ many players, are better than players at assessing a player's true abilities and therefore the optimal training regimen. In this case, training will be more effective under monopsony than under competition because the former leads to better training decisions.

Agency theory suggests another reason that the outcome of the player market may depend on player market rules. Suppose that player effectiveness is determined by another endogenous variable, player effort, e , that effort is costly, and that effort is observable to the player but not to the team. In this case, only the player directly controls effort, but the team can influence effort by picking its reward system by choosing s . The players optimization problem is to maximize net income: $sR(e) - e$. The optimal choice of e is then given by $sR' = 1$, but the socially optimal choice of effort is given by $R' = 1$, which only happens if the market is competitive. If $s < 1$ players will put forth less effort than is optimal. In the special case in which players are paid their reservation wage (monopsony is complete), players will put forth only enough effort to justify paying the reservation wage ($w = R$), but no more.

6. Nash equilibrium expenditures are zero for the team and the private optimum for the player when $s > .5$, and are the private optimum for the team and zero for the player when $s < .5$. If $s = .5$, the private optimum for players and teams is the same, and any division of this expenditure between them is a Nash equilibrium. In all cases where $0 < s < 1$, Nash equilibria are inefficient.

The preceding model provides an alternative to Rottenberg's explanation for why players are paid more than their reservation wages. Teams maximize profits by giving players an incentive to put forth effort, but because teams would lose all excess profits if they set s equal to 1, the profit-maximizing solution is to make wages responsive to player productivity, but not so responsive that players exert optimal effort. This outcome does not hinge on whether better players have bargaining power.

Coase (1937, 1960) was the first to point out that transactions costs arise from coordination problems and the inability to write complete incentive-compatible contracts. For this reason, Coase, not Rottenberg, deserves credit for the Coase Theorem and for the explosion of research in law and economics. Rottenberg's invariance hypothesis is the special case of the Coase Theorem that arises when transactions costs are unimportant. Unfortunately, the sports labor market is not an example of a market in which transactions costs can be ignored. While Rottenberg described facts that could have led him to invent transactions costs and agency theory (the role of training and the departure of salaries from reservation wages under monopsony), he did not do so.

These critiques of Rottenberg's analysis do not relegate his contribution to a mere footnote. Such a conclusion is tantamount to dismissing Newton because he did not discover relativity. Rottenberg's contribution was immense. Not only did he set the agenda for sports economics, his analysis is mostly right. Rottenberg is the father of the field, and his paper remains the most important publication in sports economics fifty years after it was published.

Neale

Like Rottenberg, Walter Neale made precisely one contribution to the economics of

sports, an essay that in some ways, as in Neale's title, is peculiar. First, the paper has a stiff humor that encourages the reader not to take it seriously. Second, notwithstanding fake footnotes to such luminaries as Ole Pro of Falstaff University and his mother-in-law, the paper has no citations, despite the fact that some of the argument duplicates Rottenberg. Nevertheless Neale's contribution has been influential because it is the first to address the economics of the organization of team sports.

Neale observed that sports is "peculiar" in that teams and athletes in individual sports, while they compete fiercely on the field, do not have an incentive to become monopolies. The reason is that the demand for sports is a demand for contests, and contests require more than one contestant. Moreover, Neale further states that a component of demand is to observe the quest for a championship – to prove over a period of time that one victory is not a fluke, but that a particular contestant is the best.

Neale draws two primary conclusions from this observation. First, the relevant unit of economic analysis in team sports – the firm – is a league, not a team, for only leagues can produce contests for a championship. Second, a league is a natural monopoly. For example, Neale argues that because the American and National Leagues in U. S. baseball meet in the World Series to determine a champion, the two leagues should be regarded as a single firm. He also regards football (soccer)⁷ and cricket to be a worldwide single firm because the former has various international competitions and the latter has test matches. Presumably if the U. S. and

7. To avoid transatlantic confusion, the term football without a modifier henceforth refers to the beautiful game. Gridiron (as it is known in Australia) is referenced as American football, with no offense to Canada intended.

Japanese major baseball leagues ever schedule a match of their respective champions, by so doing they would become a single firm.

The significance of Neale's paper is that it provides the first coherent economic argument in support of the "single entity" defense against claims of collusion against professional sports teams. A recurring policy issue in professional sports is whether teams can adopt league rules that limit competition among them in either input (players, stadiums) or output (broadcasting, team location) markets. Competition law complaints against teams claim that these practices are collusive agreements among horizontal competitors. The single-entity defense, which is set forth in detail by Roberts (1984), argues that these arrangements are not collusive because the league, not the individual team, is the firm, and that teams within a league, like two divisions of an automobile company (Chevrolet and Buick, or Mercedes and Chrysler), can not appropriately be regarded as colluding.

Neale's arguments have spawned substantial additional research, but unlike Rottenberg the consensus (but not the unanimous) view is that both of his conclusions are incorrect. Of Neale's two conclusions, the claim that a league is a natural monopoly raises the more interesting issues. Neale does not rest his argument on the obvious point that there can be only one national champion, and hence only one league. Neale's argument is stronger: direct competition among leagues is unlikely to be sustainable because competition among them for players is likely to lead to an imbalance of quality and the demise of the weaker league.

Neale believes that the argument of the owners in defending monopsonistic practices in the player market, while perhaps false with respect to teams in a league, is true for competing leagues. Neale assumes that two competing leagues will not face exactly the same demand for

games. As a result, he argues, teams in one league will acquire better players, thereby causing the other league to lose fans – and to be even more disadvantaged in acquiring players. The quality of play in the two leagues will gradually diverge until the weaker league is not viable.

The empirical support for this proposition is that in each nation precisely one major league typically is present in each professional sport. Entry by new leagues always has led either to exit by one league or to merger (usually with a loss of teams). But this observation is not compelling for three reasons.

First, the fact that leagues merge does not mean that they could not continue to exist as competitors, because merger creates additional profits if it reduces competition between the leagues. Thus, the historical willingness of governments not to impose competition policy rules against monopolization in sports is a contender for explaining the existing market structure.

Second, in one financially significant case competing leagues show no signs of collapsing to monopoly: intercollegiate sports in the U. S. At the top rung of intercollegiate sports – Division IA in American football and Division I in other sports – several different leagues operate, each crowning its own champion. Although few Division IA American football teams operate in the same city, many share regional broadcasting markets. The regional coverage of Division IA leagues is overlapping, causing them to compete for league television contracts. In Division I men's basketball, the other highly popular college sport, most large metropolitan areas have several teams, and compete for both attendance and broadcasting. Finally, many colleges engage in national recruitment of student/athletes, with top high school prospects receiving scores of inquiries and numerous scholarship offers. While intercollegiate sport exhibits substantial competitive imbalance – the top-ranked teams persist over the years

and are concentrated in a handful of conferences – the sport has not witnessed the kind of contraction that is predicted by Neale. In fact, the number of teams and leagues that play the “money” sports (American football and basketball in the top division) is growing.

Third, other institutional factors also favor the creation of monopoly leagues regardless of whether a league is a natural monopoly. For the world’s most important sport, football, the International Federation of Football Associations (FIFA) has a policy of sanctioning only one first division league in each country, of requiring that each team play only in its own national league (so leagues can not compete for teams), and of declaring all players and teams that engage in unsanctioned matches to be ineligible for international competition. Because of the financial and popular significance of the World Cup, the Champions’ League, and other regional tournaments, no national governing body can afford to risk FIFA’s sanctions by creating competitive First Division leagues or allowing a foreign team to join its top division.

In the U. S., league mergers are sufficiently suspicious under U. S. law that, since the 1960s, leagues that have sought to merge have requested a statutory antitrust exemption for a merger. In only one case – the merger of the American Football League and the National Football League – has the exemption been granted. But provisions in collective bargaining agreements that are directly related to the welfare of both parties also are exempt from antitrust. This policy enables sports leagues to avoid competition policy challenges to mergers if the terms of the merger are part of a collective bargaining agreement. The mergers of the American Basketball Association and the National Basketball Association, and of the National Hockey League and the World Hockey League, proceeded as part of a collective bargaining agreement.

If the empirical evidence is inconclusive, the theoretical argument is at best incomplete

(for a more complete argument, see Noll, 2004). To clarify the issue, consider the difference between two two-team leagues and one four-team league. In the latter case, based on Rottenberg's analysis and subsequent developments in the economic theory of sports leagues, the diminishing returns to talent and the demand for contest uncertainty lead the most advantaged team not to acquire all the talent, but instead just to be sufficiently stronger that it is the most likely winner of the championship. In the two-league case, the change in the equilibrium distribution of talent will hinge on how the four teams divide into two leagues according to market demand; however, separation into two leagues will not in general increase the demand for players for the team in the best market. For example, if the team in the best market pairs with the team in the weakest market, the effect on relative team quality (under conventional assumptions about demand) will be that the two teams will be more, not less, equal. Having said this, a loose end in the economic theory of sports league – and a very promising prospect for an important future contribution – is to characterize the effect of dividing a $2n$ -team league into two n -team leagues, using two polar cases – two leagues with equal aggregate markets, and a circumstance in which one league has the n -best markets.

Neale's other conclusion, that the league is the firm, seems without merit. The problem is *not* that Neale is incorrect in arguing that a league is a firm – it certainly can be. The error is in concluding that a team is not a firm because a league is one.

The fact that teams produce complementary inputs in the production of a contest points in a different direction – that teams and leagues are vertically related enterprises, which in some cases can become a set of teams and a league as a joint-venture cartel manager, as first argued by Sloane (1971). Consider the case of European soccer. One firm-league is the Division, such as

the Premier League. This entity schedules a double round-robin to determine its champion. The Football League is another firm-league that schedules a championship – a single-elimination tournament involving all of the teams in the four professional leagues in England. The Football Association is another firm-league which schedules a single-elimination tournament that is open to all teams, professional and amateur. And the Union of European Football Associations (UEFA) is another firm-league which organizes the Champions League (involving, among others, four Premier League teams) and the UEFA Cup (involving two more Premier League Teams). From the standpoint of economic analysis, it makes sense to regard a team in the Premier League as a member of several different leagues (three or four), selecting players and making strategic choices to maximize total profits from joint participation in all leagues.

In U. S. intercollegiate basketball, over 300 colleges field Division I teams. Nearly all (but not all) play in a league that schedules both a round-robin championship and an elimination tournament. In addition, the governing body of college sports, the National Collegiate Athletic Association, schedules two post-season single-elimination tournaments (having recently acquired its independent competitor to settle a private antitrust suit). And, numerous other organizations schedule single-elimination tournaments early in the season. Again, college basketball teams can be regarded as selling the services of their basketball teams to play games to each of these league organizers.

Teams and leagues can be vertically integrated. A league can own all of its teams, as is the case for Major League Soccer in the United States. Or, a team can create a league, recruit other teams to participate, and organize both its contests with other teams and the contests between all of the other participants. U. S. intercollegiate athletics has several examples of

basketball tournaments that are organized by one of the participating teams. Or a league can be created as a joint venture among several teams, with participation either exclusive (some U. S. professional sports teams do not play teams outside their league) or nonexclusive (U. S. intercollegiate sports, U. S. professional basketball, professional football everywhere). The existence of all organizational forms is an interesting phenomenon to be studied and explained, not an epiphenomenon to be ignored.

Jones

Unlike Neale and Rottenberg, Colin Jones has made numerous contributions to research in sports economics, and is the most productive scholar on the economics of professional hockey. The motivation for Jones (1969), his first publication in the field, was a Canadian competition policy inquiry that was motivated in part by the 1966 expansion of the National Hockey League (NHL), which rejected Canadian cities but placed six new teams in the United States. One issue that arose during this inquiry was whether hockey is a profit-oriented business, or whether team owners are motivated by the “love of the game” so that the application of competition policy to the sport is inappropriate. The main point of the article was to show that the structure and behavior of the NHL, and even its rationale for restrictive practices such as player market restrictions, are consistent with profit maximization.

Jones provides a detailed description of the institutions and rules of hockey, including the relationships between the NHL and both minor professional leagues and the association of amateur clubs. The paper analyzes the economic purposes and effects of these arrangements, and argues that their adoption is consistent with the goals of profit maximization by individual teams and joint profit maximization by the league. Many of the arguments about the underlying

economics of the product and player markets can be found in Rottenberg and a few in Neale, both of whom are cited.

Jones makes several original contributions in this paper. The first, which is ignored for different reasons by both Rottenberg and Neale, is to make a clear distinction between the theory of the league and the theory of the teams that belong to a league. As a result, part of his argument deals with the issue of how strategic behavior by a team to gain competitive advantage either is thwarted by an existing rule or creates the motivation for a change in the rules. Jones argues that in labor markets the teams of the NHL – there were only six at the time – are an oligopoly (not price-taking perfect competitors as assumed by Rottenberg) with strong incentives to cheat against collusive rules that serve the purpose of maximizing joint profits. He also argues that the league, as the cartel manager, has insufficient resources to be perfectly efficient at its task, so that the actual outcome is between the non-cooperative oligopoly equilibrium and the monopoly profit maximizing equilibrium.

Jones also provides important new insights about why players are paid more than reservation wages. Jones expands substantially on Rottenberg's analysis of the bilateral monopoly between teams and better players. His main contribution is to introduce a detailed characterization of the upper bound to the bargaining outcome: the opportunity cost to the team of failing to sign a player, which is the excess value of a player over the best available substitute.

Jones notes that a player's contribution to winning is not the only source of an opportunity cost arising from a failure to negotiate a contract. Another source arises because final demand for a sport includes the presence of a star player. Jones uses the cases of Bobby Hull and Bobby Orr, two of the greatest hockey players of all time, roughly to quantify the

independent effect of stardom. Jones also notes that group bargaining also can increase pay above reservation wages, and notes that the hockey players' union apparently did affect wages in the 1950s and 1960s.

Another new issue that is raised by Jones is league expansion. He is the first to observe the importance of broadcasting in creating an economic incentive to add teams to a league. Jones observes that in the 1960s, all NHL teams experienced growth in demand to the point that they began to reach capacity constraints. The growing popularity of hockey made the NHL potentially more attractive for national television. Jones observes that a national television contract is likely to be more lucrative if the league is national in scope. Jones argues that in its first expansion the NHL picked large U. S. markets over Vancouver and Quebec, the Canadian candidates, to maximize revenues from national broadcasting. Canadian cities would not help the NHL achieve a national television audience in the U. S. and would make a minor contribution even in Canada.

Jones' first article on sports had less impact than the other works in the original era, but not because the quality of the paper is lower. The paper suffers from its focus on hockey, which enjoys intense popularity only in Canada, and from being one of a long stream of high-quality papers by the author on the same sport, some of which address issues in this paper in a more comprehensive and sophisticated way. Nevertheless, the paper was an important contribution at the time, and remains valuable as a detailed study of the economics of the institutional structure of hockey prior to the entry of the World Hockey Association a few years later.

Sloane

Peter Sloane is the founder of sports economics in Europe. Like Jones, Sloane has

continued in the field, mostly writing about the economics of football. Sloane is the only economist to have published two papers on sports during the early period.

Sloane (1969) presents a detailed analysis of the labor market for football players that was motivated by the lack of economic content in a British government report. As with Jones, the paper covers much of the same material that Rottenberg covers, but presents much that is new. Among the new material is an excellent history of labor relations in English football, including a discussion of the watershed Eastham case, which created an arbitration process whereby, for the first time anywhere in the world, veteran English football players had the opportunity to move to another team after the expiration of their contract.

The most important analytical departure from Rottenberg is Sloane's contention that, at least for football in the United Kingdom, teams are not profit maximizers. Because this is the main issue in his companion paper, Sloane (1971), the issue of team motivation and its importance is discussed in my review of the other paper. In any event, the contribution of Sloane (1969) does not hinge on the resolution of this issue. As Sloane notes, the only important effect of firm motivation on wages is that wages in general might be lower under profit maximization; however, the basic operation of the labor market is likely to be mainly the same.

One of Sloane's contributions is to replace vague references to a reservation wage with a serious discussion of labor supply. Sloane notes that a difference in sports employment is its advantages and disadvantages beyond wages. Among the advantages are celebrity, opportunities for other earnings (endorsements, publications), and short hours. In addition, playing the game is enjoyable. Among the disadvantages are the need to avoid pleasurable activities that undermine athletic performance, a brief career, and the threat of serious debilitating injury;

however, these disadvantages are very likely to be outweighed by the advantages because many amateur athletes engage in organized sport without compensation. For all of these reasons, an athlete's "reservation wage" is likely to be less than the annual wage in the next best occupation.

Another contribution of the paper is an analysis of the economics of the "retain and transfer" system, whereby a team that employs a veteran player after the player's previous contract has expired still must pay a fee to the player's former team. The transfer system is less restrictive than the player reservation system that was the focus of Rottenberg's and Sloan's analysis. The transfer fee system is a form of the "option-compensation" system.⁸ As Sloane states, the transfer system is the linchpin of the football labor market – or it was until it was undone by the Bosman decision by the European Court of Justice in 1996.

Sloane analyzes the effects of player market monopsony, and concludes that the transfer system reduces wages; however, he disagrees with the conclusion that, therefore, players are exploited, using as the definition of exploitation whether wages are less than the marginal revenue product.⁹ The reason Sloane gives is that if teams are not profit-oriented and have a

8. Under the option-compensation system, a player can become a "free agent" by playing one year after a contract expires at the salary paid in the previous year. If the player signs with another team, the player's old team is compensated. Whether this system gives players a realistic opportunity to move depends on the compensation process. In American football (the National Football League) until the late 1980s, the Commissioner of the sport decided the amount of compensation, with the magnitude unlimited and unchallengeable.

9. The term "exploited" carries baggage, and I use it only because it is commonly used in labor economics in the manner that it is used here. I believe that the better approach is to refer to the

propensity to overpay players, the transfer system may do no more than prevent excess wages.

Sloane notes that over half of English football teams run a loss on football operations, from which he concludes that despite the transfer system the wage, on average, must exceed the marginal revenue product of the player, so that players are not exploited.

This conclusion is incorrect as a general proposition of economic theory, even if one accepts the dubious proposition that accounting profits are a valid indicator of economic profits. As a technical matter, profitability refers to the total conditions, not the marginal conditions, of optimizing firm behavior. To see this, construct any example of a firm that enjoys a product market monopoly and acquires inputs in competitive markets, using any arbitrary choice of demand and cost functions that produces an equilibrium in price, output and inputs. Once the example is in hand, calculate the equilibrium profit, M , and consumer surplus, S , then add a fixed cost to the cost function equal to $M + m$, where $0 < m < S$. The new equilibrium will be identical to the old in price, output and inputs, the firm will create positive social value, inputs will be paid their marginal revenue products, but the firm will be unprofitable. Could such a model explain the conditions of a sports team, at least in the short run? Of course. In the short run (a season), many costs of the team are fixed, such as the playing facility, the costs of traveling to away games, the minimum staff necessary to field a team, and other contractual agreements with players, coaches and executives.

Sloane's inference about exploitation is incorrect for another reason. In England, where

efficiency of the labor market, in which case the appropriate standard is whether the wage equals the value of the marginal product. The discussion here, however, adopts the conventional terminology and standard.

teams do not have territorial exclusivity and the largest cities host several professional teams, a plausible candidate for the most accurate theoretical model of a team is product differentiated (monopolistic) competition, where teams are differentiated according to the league in which they play and their relative quality within a league. In this case expected profits are near zero.¹⁰

Owing to random factors that affect performance (such as injury or unanticipated changes in the skills of players), in any year annual profits should be at or below zero for around half the teams.

Another possibility is that the relevant differentiated product model is not monopolistic competition, but either oligopoly or a dominant firm with a competitive fringe. Such a model is relevant if, as an empirical matter, substitutes (say, Manchester City) are inherently inferior and so can not compete away the excess profits of the most popular teams (say, Manchester United). In this case players for the latter club may be exploited even if players for the former are not.

Finally, Sloane does not use the appropriate measure of marginal revenue product to detect exploitation if, as he argues, teams are not profit maximizers. Sloane argues that team owners are utility maximizers. If so, the conventional measure of marginal revenue product understates its actual value. A team's demand for players is derived from the willingness to pay

10. Expected profits may be positive because the Football League limits the total number of professional teams to 92; however, entry and exit into the bottom league through promotion and relegation with the highest semi-professional league allows the geographic distribution of teams to adjust, albeit slowly, to changes in relative demand across localities. In any case, the largest cities all have multiple teams in the top two leagues, and if there were substantial excess demand in these markets for still more teams in top leagues, teams from these localities that play in lesser leagues ought to be able to generate enough revenue to sustain them in a higher league.

of fans for team quality, which in turn is derived from the utility maximizing behavior of fans. In the simplest case of utility-maximizing teams, the owner maximizes a function of team quality and other consumption, subject to a budget constraint that includes revenues from team operations and personal wealth. In this case, the derived demand for players is the sum of the marginal willingness to pay across both fans and owners. Put another way, in the utility-maximizing model, an owner is also a fan, and if an intense fan becomes an owner, that person's demand for team quality should not suddenly be excluded from either the positive or the normative model of team operation. If one believes, as does Sloane, that the objective function of a team includes team quality, then the test for exploitation is not whether wages equal marginal revenue product as conventionally calculated, but whether they equal conventional marginal revenue product plus the marginal willingness to pay of the owners. If the wage equals the marginal revenue product as derived from a demand curve that excludes the willingness to pay of the team owner, players are exploited because the wage is less than the marginal increment to social welfare that the player creates.

An innovation in Sloane (1969) is his analysis of the economics of transfer fees. He presents a model of the appropriate method of calculating the maximum transfer fee a team should be willing to pay, based on the discounted present value of the additional net revenues that the player will generate. Sloane correctly observes that this calculation should take into account the tax consequences of the acquisition, which allows the team to use the transfer fee as a deductible expense (rather than a capital expenditure that is subject to amortization over the period that the player plays for the team). He then provides an example that translates a transfer fee for a player with a remaining playing life of five years into the increment to attendance that

the player must generate to justify the acquisition. Although one can quibble over the details of the example, this calculation remains the basic model that students are taught today in evaluating both transfer fees and signing bonuses.

Sloane (1971) remains one of the most important papers in the sports economics. Although its most influential contribution is its development of an economic theory of a sports team as a utility maximizer, there is much of value in the paper other than this. Whereas Sloane (1969) contains the history of player relations, Sloane (1971) focuses on the organizational structure of English football, including revenue sharing and allocating revenues from tournaments. This paper is the first to present an economic rationale for revenue sharing – why it is not necessarily an inefficient form of cross-subsidization. The analysis builds on the observation of Rottenberg and Neale that teams are complementary inputs to the production of games and championships. Sloane adds the point that the optimal structure of a league can be to include some firms that are not viable. In essence, the weaker teams may enhance the demand for the stronger teams such that the revenue enhancing effect on the latter exceeds the financial losses of the former. This argument dovetails nicely with Jones' analysis of the expansion of the NHL to create a market for national broadcasting rights.

Notwithstanding these other valuable contributions, Sloane (1971) is still cited frequently because it presents a model that conforms to the belief among many European economists that teams do not maximize profits. No controversy in the economics of sports is more intense than the debate over the business objective of sports teams. The significance of this debate is the widespread belief that the efficiency of various forms of collusion among teams and leagues, and hence the most appropriate competition and labor policies in sports, hinge on the motivation of

team owners.

The current state of this debate is an intellectually unsatisfying truce. Among sports economists, the profit-maximization hypothesis is generally regarded as applicable for most teams in North America and among the top teams in European football, whereas the utility-maximization model is generally regarded as appropriate for other teams in Europe. This accommodation, while making conferences of sports economists more pleasant, is intellectually unsatisfying because no coherent explanation has been offered for why the economic theory of sports should depend on which side of the Atlantic Ocean a league is located. The reigning explanation is that, for the most part, European football teams emerged from multi-purpose local clubs whose members, because they are mostly avid fans, seek to win rather than to be wealthy.

This account, though it does not necessarily imply that clubs do not maximize profits (as explained below), needs further elaboration in the contemporary world in which football teams take in hundreds of millions of dollars and can have capitalized values of as much as a billion dollars. These numbers imply an enormous shadow price on wins for utility-maximizing owners. Or, if the better teams in the top divisions are profit maximizers but the rest are not, how and when did the transformation of objectives occur among those at the top? And when a team ascends from the ranks of the utility maximizers to the rarified atmosphere of profit maximizers, can we detect the change in motivation? These are gaping holes that call into question the validity of the truce among sports economists.

In developing an alternative theoretical model of a team, Sloane assumes a general, and therefore quite complicated, team utility function. The arguments include the success of the team, team attendance, the intensity of competition in the league as a whole, and profits. The

budget constraint depends on operating profits, subsidies from other teams or governments, and ongoing capital contributions from owners if the team persistently is unprofitable.

Sloane does not derive equations that express the equilibrium conditions for a team, let alone develop a general equilibrium model of a league in which firms behave as Sloane assumes. The main point of Sloane's formulation is to emphasize the possibility of a trade-off between profits and other objectives, notably the quality of the team and the quality of the league. Amazingly, a formal model of a league of utility-maximizing teams was not developed until thirty years later by Fort and Quirk (2004), who unfortunately forgot to cite Sloane (1971) and in so doing missed an opportunity to compare their results with his conjectures.

Sloane argues that under utility maximization, but not under profit maximization, owners of a team in a small market will not be satisfied with a perpetual state in which the expected outcome for a season is for their team to finish out of the running. As a result, contrary to Rottenberg's analysis, teams that are "too good" for the market in which they operate will not sell their best players to teams that are "too bad" for their market, even if the transfer fee is more than sufficient to compensate them for the lost revenues from the transaction. If player markets are unrestricted, and players either maximize income or maximize some combination of income and team success, player movements to teams in better markets are not inhibited by the poorer team's utility-maximizing behavior. Under utility maximization, diminishing returns to team quality will not create as much of a constraint on team behavior, as good teams will sacrifice profits to become even better.¹¹ Thus, leagues of profit-maximizing teams will be larger, and if

11. Sloane's argument is incomplete. At the profit-maximizing equilibrium level of team talent, the marginal utility of league health could exceed the marginal profitability of the last player

teams maximize utility a “death spiral” – good teams acquire good players and poor teams become uncompetitive and drop out of the league – is more likely.

The mechanics of how a death spiral might arise are clear from the results of Fort and Quirk (2004), who show that a league of a fixed number of utility maximizing teams has a higher equilibrium price for talent and acquires more talent than a league of profit-maximizing teams. Greater expenditures on talent by teams in better markets could make teams in weaker markets financially unviable regardless of their objectives, so their results are broadly consistent with Sloane’s argument.

Sloane derives two welfare predictions from his analysis. The first is that under the utility maximization hypothesis and an unrestricted market for players, professional leagues will be too small; however, he notes that many teams apply to join the English Football League whenever an opening arises, so that an easing of restrictions might not cause the Football League to shrink.¹² The second is that the distribution of team quality is more dispersed, which in turn reduces consumer welfare. This loss arises not just from the fact that bad teams are worse than they would be under profit maximization, but that good teams will be too good, causing loss of

plus the marginal utility of wins. If so, leagues of utility maximizing teams could be more competitively balanced than a league of profit-maximizers. Sloane implicitly assumes that this circumstance is implausible, in which case the model can be simplified without sacrificing the qualitative results by dropping league health as an argument of the utility function.

12. At the time Sloane wrote his paper, promotion and relegation between the Conference and the League was not automatic as it is now, but arose from an applications process in which applicants were compared with the worst performers in the League.

welfare by their fans as well due to the loss of excitement because matches are not as close.

Accepting the assumptions that lead to Sloane's conclusions about the effect of utility maximization on the distribution of talent, his welfare conclusion is incorrect. The reason is that welfare is derived from utility, and his analysis ignores the utility of owners. Sloane (1971) makes the same omission in aggregating preferences that is discussed in the review of Sloane (1969). The welfare of a fan should not be ignored just because the fan becomes an owner. Moreover, to upset the trans-Atlantic truce that allows both theoretical models to survive in the literature, I assert that the predictions of the utility-maximizing model are the same as the predictions of a profit-maximizing model that takes into account why utility-maximizing teams are formed and how they behave. (I hope I remain welcome at conferences.)

Recall Sloane's assumption that a crucial difference between the two types of teams is that utility-maximizers make capital contributions to support the losses of the team. This assumption raises another question: under these conditions, why would a rational person become an owner of a utility-maximizing team? The answer must be more than that an owner derives utility from ownership. If owners, like players, derive utility from their occupation, the psychic income from ownership will not cause the team's behavior to differ from standard profit maximization. In this case, the owner settles for less than the competitive return on investment, pays too much for the team in relation to its cash flow, and is unwilling to sell at a price that reflects its value as a financial asset. Nevertheless, the owner would still want to minimize the cost of acquiring this psychic income, which can be accomplished by maximizing team profits because there is no trade-off between profits and utility from ownership.

To add bite to the utility-maximizing model of ownership requires assuming, as does

Sloane, that owners are people who, at current ticket prices, are willing to pay more in order to have the team have a better record. In short, they are fans whose demand for winning at current prices is less elastic than the average market elasticity of demand for quality among a team's supporters. By buying a team, a coalition of fans can translate their atypically low elasticity of demand for team quality by engaging in price discrimination against themselves. Club ownership, then, is a method for identifying fans with inelastic demand, and the capital contribution of these owners is a form of price discrimination.

Under these assumptions, fan-owners seek to maximize their own surplus, which is profits plus the maximum willingness to pay of owners for team quality less the amount that they actually pay, including ticket prices and capital contributions. The capital contribution is analogous to the first part of a two-part tariff in that capital contributions are not a tax on the marginal unit of consumption, such as an increase in ticket prices or concession goods. Thus, the capital contribution does not create dead-weight loss in the consumption of contests.

To explore the properties of this model, temporarily assume that profits are distributed equally across ownership shares with no further distortions (which is not true). The problem of utility maximization for owners then has multiple equilibria. At the level of team quality that would maximize team profit (including capital contributions as a head tax), any per capita contribution between the amount that creates zero profit and the amount that maximizes team accounting profits is equally attractive. In reality the distortion from taxation will cause one equilibrium to be selected. In particular, teams are likely to be subject to an income tax or to face limits on distributions to owners, in which case the optimal contribution is set so that the team just breaks even. Nonetheless, the quality of the fan-owner team is identical to the quality

of a profit-maximizing team that can practice the same price discrimination against rabid fans.

Because the ownership contribution is not really a head tax, distortions creep in. Rabid fans are not likely to be equally rabid, in which case distortions arise because, as the capital contribution rises, fewer rabid fans want to be owners. The solution to this problem is to make the team a joint stock company, enabling owners to pick the number of shares that they own, so that less rabid fans can buy less ownership and make lower capital contributions. Public policies also affect the choice of organizational structure and the amount of owner contributions. For example, the joint stock option faces more stringent regulatory rules than the partnership or non-profit club option, in which case the cost of avoiding distortions arising from differences in preference intensities among owners may exceed the benefits.

This analysis does not necessarily lead to the conclusion that the dominant organizational form is one in which rabid fans are owners. To complete the analysis, imagine that a profit-oriented owner has decided to capture additional profits from price discrimination against rabid fans. This owner can pursue two strategies. One is to sell the team to rabid fans who then can make capital contributions. The other is to create a membership organization for collecting fees beyond ticket sales, recognizing that the demand for membership will depend on team quality. The choice between these strategies depends on which option produces a bigger gain for the owner, which in turn depends on the transactions costs of alternative organizational forms.

If a profit-oriented owner decides to implement price discrimination, some sources of transactions cost are to identify fans who are profitable targets of price discrimination, to create a way to implement the policy, and to pick the optimal price. The owner must create a substitute for ownership for rabid fans, a form of special club with perquisites, for which a fee can be

charged. Price discrimination of this sort is found not only in sports but also the performing arts. Examples are personal seat licenses and booster clubs.

Another transactions cost is the potential for dead-weight loss in picking the price. A profit-oriented owner has no direct analog to the joint stock company of fan-owners, so that the non-ownership solution to the price discrimination problem may be more distorting than the optimal fan-owner solution. In this case, team quality and hence fan utility under a price discriminating profit-oriented owner will be less than under a joint-stock fan-owner arrangement.

The fan-owner option also suffers from transactions costs. One is the collective action problem: rabid fans who do not buy ownership shares and make capital contributions enjoy the same improvement to team quality as the rabid fans who do buy shares, thereby creating an incentive to free ride. Another problem is that fan-owners must determine the capital contributions among them. Typically these decisions are made through voting, which produces the preferred choice of the median voter rather than the choice that is the profit-maximizing fee and team quality. Still another problem is that rabid fans may not make the best decision makers when exercising their managerial prerogatives, so that the team may be managed less efficiently.

For all of these reasons, the optimal ownership structure of a team is not obvious. In fact, the optimal organizational form may differ among teams depending on exogenous market factors that affect transactions costs, such as the nature of the team's fan base and the likely prospects for positive accounting profits in the long run. Thus, the choice of organizational structure by teams appears to be an extremely interesting object of research, but one cost of the transatlantic truce is that this issue has been ignored.

The welfare implications of a league in which teams practice greater price discrimination

(and hence of fan-owner teams) are indeterminate, but are not alarming. The conjecture by Sloane, proved by Fort and Quirk, that players earn higher salaries under utility-maximizing owners remains correct, but the model outlined here provides a clearer picture of why. Price discrimination against fans causes an outward shift in the derived demand for players, which drives up the equilibrium wage if the supply of players is not perfectly elastic.

Also as shown by Fort and Quirk, if the supply of talent is not perfectly inelastic, the outward shift in demand leads all teams to acquire more talent and thereby increases the absolute quality of play. The new equilibrium benefits fans who are not owners. The quality of play is a public good among all fans of a sport, and any form of effective price discrimination against rabid fans will deliver more quality to other fans. In addition, a testable theoretical implication of the model is that price discrimination against rabid fans leads to lower ticket prices (per unit quality) and hence greater welfare for less rabid fans. Some non-rabid fans nevertheless could be harmed if the relative quality of their team were adversely affected. Whether this will be the case depends on whether the distribution of rabid fans among teams differs substantially from the distribution of non-rabid fans. Under price discrimination, relative quality increases for teams with proportionately more rabid fans.

A general movement to league-wide price discrimination is likely to harm rabid fans. If the distribution of rabid fans among teams is roughly the same as the distribution of non-rabid fans, price discrimination leaves the relative quality of teams unchanged, so rabid fans do not succeed in their attempt to increase their utility by winning more matches. Yet rabid fans, through their capital contributions, finance the increase in player salaries and bear more of the other costs of the team. Thus, unless rabid fans place a sufficiently high value on absolute team

quality, the new regime harms them.

The most important implication of this analysis is that the motivation of team owners is irrelevant from the perspective of competition and labor policy. If the fan-owner utility-maximizing model is equivalent to a profit-maximizing model in which teams engage in price discrimination against rabid fans, there is no valid reason to treat fan-owners differently with respect to the legal status of their behavior in either input or output markets. Doing so will reduce efficiency in two ways. First, it will induce profit-oriented owners to sell to fan-owner groups because the latter will be entitled to greater implicit profit that arises from the redistribution effect of monopsony. The latter effect can cause fan-owner teams to replace profit-oriented owners even if the latter are more efficient. Second, giving special privileges to fan-owners in comparison with profit-oriented owners will increase distortions in markets in which fan-owner teams are permitted to act anticompetitively.

El-Hodiri and Quirk

Mohammed El-Hodiri and James Quirk occupy a unique niche in the intellectual history of sports economics. El-Hodiri and Quirk (1971) developed the first mathematical model of a sports league that is built from independent profit-maximization by member teams. Their paper has two motivations. As with the other papers in this period, their article is motivated by policy concerns – whether sports leagues should have an antitrust exemption. In addition, as appreciative readers of Rottenberg, they seek to state with precision the conditions under which the invariance hypothesis is true, and whether institutional rules other than the player market restrictions then in place could affect the distribution of playing talent among teams.

The formal model assumes that teams maximize profits, and examines the equilibrium

profits and qualities of teams in an n-team league in which relative team quality is determined by the distribution of player skills among teams. The model examines three forms of player markets: complete free agency, no free agency but sales of players between teams, and neither free agency nor sales of contracts. The model is based on three further core assumptions, all of which are controversial and have been altered in subsequent work in the field.

The first core assumption is that the supply of player skills is fixed. In the model each team is endowed with a stock of depreciating player skills and an annual flow of new player skills (such as through a draft). This assumption corresponds to a league in which the only source of new players is novices who have no alternative equivalent employment. As such, it fairly reflects conditions in Canada and the U. S. in periods when all the major North American sports were not played at the highest level anywhere else, all the leagues were monopolies, and the rookie drafts in all leagues gave each team enough draft choices such that virtually all players with a non-trivial chance of becoming a major league player were drafted. This assumption does not apply if several leagues operate at the highest level of play, such as is the case for football. Ironically, the assumptions did not fully apply even in the U. S. at this time because of the entry of competitive leagues in every sport except baseball.

The second core assumption is that the revenue function is concave in relative team quality. In words, this assumption implies diminishing returns to quality, and is a mathematical representation of the claim in earlier papers that fans prefer uncertainty of outcome. In practice, this assumption is slightly weaker than the statements in some of the earlier papers that certainty of outcome leads to a decline in revenues. All that this assumption requires is that the increase in revenues due to an increase in quality diminishes as quality increases. The controversy here

arises from the fact that in some cases the winner may receive a prize for being the best team. An example is that the top four teams in the Premier League qualify for the Champions League, causing a huge increase in revenues if a team moves from fifth to fourth place. In this case the revenue function is not concave in team quality over a range of quality that represents the best teams in the league.

The third core assumption is about the identification of the equilibrium. El-Hodiri and Quirk assume that the wages and contract prices per unit of skill are the same for every team, that each team assumes that it has no effect on wages, the price of player contracts, or the distribution of playing skills among other teams, and that each team makes a decision about whether to acquire playing skills under the assumption that no other transactions will occur. These assumptions rule out strategic behavior or market power.

From these assumptions, El-Hodiri and Quirk prove four main results. First, if there is a market for either players or player contracts, equal distribution of playing talent arises only if all teams face the same revenue function, which implies that all local markets in which team operate are equally attractive. Second, in the presence of revenue sharing, equal distribution of talent requires that the share of revenues kept by the home team exceeds 50 percent. Third, if markets do not have the same revenue potential, equal playing strengths (competitive balance) can be maintained only if leagues ban the sale of player contracts as well as competition for players. Fourth, the first and third results remain true if players must be trained to acquire skills.

The 1971 paper did not address the invariance hypothesis, but sought only to identify conditions under which equal playing strength was an equilibrium. In Quirk and El-Hodiri (1974), the authors extend the model to determine the equilibrium distribution of playing

strengths when markets are not equally attractive. The 1974 paper shows that teams in better markets have higher quality, that the distribution of playing strengths is not affected by revenue sharing, that player salaries are increasing in the share of revenue that is kept by the home team, and that the distribution of playing strengths when contracts can be bought and sold is the same as the distribution that would arise under a cartel that maximized joint profits. This version of the invariance hypothesis is stronger than postulated by Rottenberg, for among the institutional features that do not affect the allocation of players are the revenue sharing rules and whether teams are independent companies or the firm-league hypothesized by Neale.

El-Hodiri and Quirk (1971) conclude by returning to the issue of an antitrust exemption. They observe (p. 1319) that the baseball antitrust exemption has not led to the adoption of rules that would cause more equal competitive balance; however, they also state that an antitrust exemption that would preserve player market monopsony is “amply justified” if it is accompanied by a ban on sales and trades of player contracts. The basis for this proposal is “externalities peculiar to professional sports.” This conclusion has no basis in the model that is presented in the paper, for the justification, if it exists, depends on the nature and magnitude of these externalities as well as the perpetuation (and implicit antitrust exemption) of exclusive territorial franchises that inhibit the equalization of team revenue functions.

Notwithstanding the gaffe in the final paragraph, El-Hodiri and Quirk made a seminal contribution to the economics of sports by establishing the theoretical framework for examining the properties of institutional rules in professional sports. The paper generated an extensive literature that seeks to examine the implications of different assumptions within the framework of profit maximizing teams and in particular conditions that undo the invariance hypothesis.

Although others have succeeded in this quest, the original contribution remains the gold standard by which alternative models are evaluated.

The Legacy

By the end of 1971, the six original pieces had built a firm foundation for creating the field of sports economics, but there were some conspicuous gaps. The most obvious is the absence of more than casual empirical research. The original six papers do not contain a single regression. This imbalance no longer exists, but the field still has unresolved controversies about how best to measure player quality, team quality, and the uncertainty of outcomes.

In addition, because the early papers were motivated primarily by issues in competition and labor market policies, they do not address, or address only incompletely, many equally important issues. Examples are the economic impact of sports in a community, the market for team locations, public subsidies of sports, sports broadcasting, the organization of teams and leagues (as is emphasized in my critique of the early works), labor supply (including the relationship of professional to amateur sports), and sporting events that do not fit neatly into the category of leagues of professional teams (including intercollegiate athletics in the U. S. and international mega-events, like the World Cup and the Olympics).

The existence of gaps is not a criticism of the early works, for these six papers accomplished a great deal. These papers established an enduring conceptual framework for analyzing the economics of sports. In the process, these papers pretty much destroyed much conventional mythology about the economics of sports, for which the authors deserve great praise. Thankfully these authors gave us a foundation on which to build, but left the rest of us something to do.

References

- Andreano, Ralph (1965). *No Joy in Mudville: The Dilemma of Major League Baseball*. Cambridge, Massachusetts: Schenkman.
- Andreff, Wladimir, and Jean-Francois Nys (1986; 5th edition 2002). *Economie du Sport*. Paris: University Presses of France.
- Bain, G.S. (1967) *Trade Union Growth and Recognition*, Royal Commission on Trade Unions and Employers Associations Research Papers no. 6, London: HM Stationary Office.
- Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.
- Jean-Francois Bourg and Jean-Jacques Gouget (1998). *Analyse economique du Sport*. Paris: University Presses of France.
- Buch, Martin Peter, and Horst-Manfred Schellhaas (2005). *Okonomik von Sportligen*. Schorndorf, Germany: Verlag Karl Hofmann.
- Coase, Ronald (1937). “The Nature of the Firm.” *Economica* 4 (November), 386-405.
- _____. (1960). “The Problem of Social Cost.” *Journal of Law and Economics* 3 (April), 1-44.
- Frank, Robert H., and Philip J. Cook (1996). *The Winner Take All Society*. New York: Penguin.
- Downward, Paul, and Alister Dawson (2001). *The Economics of Professional Team Sports*. London: Routledge.
- El-Hodiri, Mohammed, and James P. Quirk (1971). “An Economic Model of a

Professional Sports League.” *Journal of Political Economy* 79 (November-December), 1302-19.

Fort, Rodney (2003). *Sports Economics*. Upper Saddle River, N. J.: Prentice-Hall.

_____(2005). “The Golden Anniversary of ‘The Baseball Players’ Labor Market.””

Journal of Sports Economics 6 (November), 347-58.

_____, and James Quirk (2004). “Owner Objectives and Competitive Balance.”

Journal of Sports Economics 5 (February), 20-32.

Gitelman, H. M. (1968). “An Investment Theory of Wages.” *Industrial and Labor Relations Review* 21 (April), 323-352.

Heinemann, Klaus (1995). *Einführung in die Okonomie des Sports*. Schorndorf, Germany: Verlag Karl Hofmann.

Jones, J. C. H. (1969). “The Economics of the National Hockey League.” *Canadian Journal of Economics* 2 (February), 1-20.

Kirchhoff, Sue (2006). “Batter Up! Sports Economics Hits Field.” *USA Today* (July 27), available at http://www.usatoday.com/money/2006-07-27-sports-econ-usat_x.htm.

Leeds, Michael, and Peter von Allmen (2005). *The Economics of Sports*. Boston: Addison Wesley.

Li, Ming, Susan Hofacre, and Dan Mahoney (2001). *Economics of Sports*. Morgantown, West Virginia: Fitness Information Technology.

Neale, Walter C. (1964). “The Peculiar Economics of Professional Sports.” *Quarterly Journal of Economics* 78 (February), 1-14.

Noll, Roger G. ed. (1974). *Government and the Sports Business*. Washington: Brookings Institution.

- _____. (2003). "The Organization of Sports Leagues." *Oxford Review of Economic Policy* 19 (Winter), 530-51.
- Quirk, James P., and Mohammed El-Hodiri (1974). "The Economic Theory of a Professional Sports League." In Roger G. Noll, ed., *Government and the Sports Business*. Washington: Brookings Institution.
- Ritter, Lawrence S. (1966). *The Glory of Their Times*. New York: Macmillan.
- Roberts, Gary (1984). "Sports Leagues and the Sherman Act." *UCLA Law Review* 32 (October), 219-301.
- Rottenberg, Simon (1956). "The Baseball Players' Labor Market." *Journal of Political Economy* 64 (June), 242-58.
- Sanderson, Allen R., and John J. Siegfried (2006). "Simon Rottenberg and Baseball, Then and Now: A Fiftieth Anniversary Perspective." *Journal of Political Economy* 114 (June), 594-605.
- Sandy, Robert, Peter J. Sloane and Mark S. Rosentraub (2004). *The Economics of Sports*. London: Macmillan.
- Sloane, Peter J (1969). "The Labour Market in Professional Football." *British Journal of Industrial Relations* 7(2): 181-199.
- _____. (1971). "The Economics of Professional Football: The Football Club as a Utility Maximiser." *Scottish Journal of Political Economy*, 18 (June), 121-46.
- _____. (2006). "Rottenberg and the Economics of Sport after 50 Years." IZA Discussion Paper No. 2175 (June). Bonn: Institute for the Study of Labor.
- Trosien, Gerhard (2003). *Sportökonomie*. Aachen, Germany: Meyer and Meyer.

U. S. Bureau of the Census (2004a). "Museums, Historical Sites, and Similar Institutions: 2002." Washington, D. C.: U.S. Department of Commerce (July).

_____ (2004b). "Performing Arts, Spectator Sports, and Related Industries: 2002." Washington, D. C.: U.S. Department of Commerce (August).

_____ (2004c). "Fruit and Vegetable Canning: 2002." Washington, D. C.: U.S. Department of Commerce (December).

_____ (2004d). "Dog and Cat Food Manufacturing: 2002." Washington, D. C.: U.S. Department of Commerce (December).

_____ (2004e). "Other Animal Food Manufacturing: 2002." Washington, D. C.: U. S. Department of Commerce (December).

U. S. Bureau of Economic Analysis (2006). "Gross-Domestic-Product-by-Industry Accounts," *Industry Economic Accounts*. Washington, D. C.: U. S. Department of Commerce (October 5), at http://bea.gov/bea/industry/gpotables/gpo_action.cfm?anon=243&table_id=17292&format_type=0.

Weintraub, Sidney. 1958. *An Approach to the Theory of Income Distribution*. Philadelphia: Chilton.