Economics of Intellectual Property Protection for Software: The Proper Role for Copyright

Frederick R. Warren-Boulton,
Kenneth C. Baseman,
MiCRA: Microeconomics Consulting and Research Associates, Washington, D.C.
and Glenn A. Woroch
University of California, Berkeley

Economics of Intellectual Property Protection for Software: The Proper Role for Copyright

Frederick R. Warren-Boulton, Kenneth C. Baseman,
MiCRA: Microeconomics Consulting and Research Associates, Washington, D.C.
and Glenn A. Woroch
University of California, Berkeley

■ This article presents an economic analysis of the proper scope of copyright protection for computer software. We begin by identifying key economic characteristics in the production and use of software; notably, the cost to developers is largely fixed and sunk, users often incur substantial sunk costs, and the value of software to users is usually a significantly increasing function of the total number of users (i.e., "network externalities" are important).

We then use economic theory and analysis to establish three propositions. First, we demonstrate that the copyright protection granted to the original developer of a software package should not extend to elements of the software that achieve the status of a *de facto* standard because the resulting monopoly leads to pricing that fails to achieve efficient dissemination of the software and fails to reward other sponsors who have invested in the *de facto* standard. Next, we argue that software interface specifications also should not be copyrightable since that would permit an inefficient extension of market power to complementary software and to later improvements. Finally, we favor reverse engineering for the purpose of achieving interoperability, since it enables firms to efficiently design compatible programs and to guard against unwarranted abuse of copyright protection.

In most instances, recent case law is consistent with these principles, especially since the recent Appeals Court decision in *Lotus v. Borland.* Significantly, copyright law has devised a "merger doctrine" that denies protection whenever a product is the (nearly) unique expression of an uncopyrightable idea, a principle that effectively implements our prescriptions for software copyright. Since we conclude that copyright is the appropriate form of protection for intellectual property only when the likelihood of an unwarranted grant of monopoly is extremely low, this prescription achieves the desired balance between the need to reward developers of innovative software programs and the need to encourage suppliers of complementary products and those who build upon and advance prior work.

Determining the optimal form and degree of protection for intellectual property poses special problems. Private sale of information has several inherent difficulties. In particular, it is difficult to establish the value of information without at the same time revealing that information, at which point the potential buyer has acquired the information at no cost.[1] Governmental intervention is clearly desirable to establish property rights in information and to prevent users from "free-riding" on the efforts of its creators. In attempting to determine the optimal extent of those property rights, however, the policy maker faces a dilemma caused by another peculiar feature of intellectual property: it displays the characteristic of a "public good" in that it is costly to produce but costless to use. Consequently, efficiency in production requires that the producer receives a positive price, whereas efficiency in distribution requires that users should pay a zero price.

It comes as no surprise, therefore, that reasonable observers differ vastly in their views regarding the appropriate level of protection for intellectual property. Those who focus on the free-rider problem allege that producers are often unable to secure even a fraction of the value of their work. Efforts to limit protection are typically perceived simply as attempts by less able competitors to handicap their more ingenious rivals, leading to the conclusion that intellectual property is unlikely ever to be overprotected in practice.[2]

On the other side of the issue are those who contend that few discoveries are made in isolation, and that most scientific advances build on the contributions of others, many of which are in the public domain.[3] A breakthrough in basic research in a related field, or spillovers from more distant technical areas, can greatly reduce the cost of developing specific

1 See Arrow [1962].
2 For example, see Landes and Posner [1989]; Miller [1993]; and Judge Keeton's decision in *Lotus v. Paperback.*
3 See Barzel [1968].

Our criterion for overprotection is whether [such protection] leaves the rest of society worse off

applications. Allowing the developer exclusive rights to those marketable applications allows the grantee to profit from prior work in excess of the grantee's actual contribution. Beyond any equity considerations, such overprotection is inefficient; it results in excessively high prices that in turn cause underutiliziation of information.[4] Equally important, it can lead to wasteful competition over rights that dissipates much of the value of the underlying prior work. This occurs when potential grantees devote excessive resources to winning a "patent race."

This article examines one particular kind of intellectual property protection: copyright protection for computer software. Our analysis does not attempt to determine the "optimal" level of such protection,[5] but instead seeks appropriate boundaries for that protection. Our criterion for overprotection is whether copyright protection of a piece of software leaves the rest of society worse off; i.e., whether the copyright holder receives more in monopoly rents than the copyrighted software has added to total welfare.[6]

Overprotection would be possible when, perhaps because of prior work, the discovery would have been made soon anyway without any protection whatsoever, or with much less protection, and thus at lower cost to the rest of society. Intellectual property protection would also be excessive if the grantee could raise the costs and reduce the opportunities available to suppliers of substitute products.

Copyright law attempts to prevent such leveraging, however, by distinguishing between "idea" and "expression." An expression can operationally be defined as a discovery or work that would not otherwise have been made (e.g., no one besides Mary Shelley would have written *Frankenstein*), and protection which would not reduce the opportunities available to any other author. In contrast, granting exclusive rights to an idea (e.g., a novel about the creation of a humanoid monster) would significantly limit alternatives available to other authors and would allow the copyright holder to appropriate the value of something she did not create.

To determine the efficient bounds on the application of software copyright, however, we need to introduce two concepts from economics. The first is the notion of "network externalities." Network externalities occur when the value of a product or service increases with the cumulative number of users. When this is the case, each additional purchase raises the value to existing users as well as the expected value to future adopters.

The second economic concept follows from a simple expected-value calculation.[7] We argue that the uncritical, unexamined, and extensive nature of protection under copyright is efficient only when the expected value of the welfare lost by protection being granted mistakenly is *de minimis*. Like summary judgment and the *per se* rule, copyright is the cost-effective approach only when the probability of error (i.e., overprotection) is very small, and/or when the potential social loss associated with error is very small.

As we argue below, allowing the copyright of interface specifications or of a program's elements that become *de facto* standards can violate our criterion for overprotection (i.e., can make the rest of society worse off) when network externalities are present.[8] In that event, the copyright holder may be able to appropriate the result of the efforts of others and to raise the costs or reduce the opportunities available to suppliers of substitute products.

A similar concern arises where copyright of interface specifications would enable the copyright holder to control the supply of compatible complementary products. At best, the welfare effects of allowing such "vertical control" are ambiguous, implying that at least some justification should be offered before the state grants property rights that would permit such control.

This does not necessarily mean that innovators should be denied rights to software that has these properties. Where the innovator has made a substantial contribution, property rights may be available through patent or trademark protection. Interface specifications and standard program elements should not, however, receive the extensive protection that is automatically and uncritically conferred by copyright.

4 Unless, of course, the seller was able to perfectly price-discriminate among users, an impossible task given the asymmetry of information.

5 Such an exercise would have to address the "public-good" nature of information. Specially, it would require some practical way of ensuring that producers of information recover an amount only slightly in excess of its cost, assuming that cost is less than the social value of the information. This rule has been proposed by Landes and Posner [1989] and Menell [1987, 1989].

6 More precisely, the *ex ante* expected value of the reward to the winner of a copyright should be related to the *ex ante* expected value of the winner's incremental contribution over and above the surplus that would have been generated if the other contestants had competed in the absence of the actual winner.

7 "Expected value" is the sum of the probability of each outcome multiplied by the value of that outcome. For example, the expected value of a 1/4 chance of $10 and a 3/4 chance of $20 is equal to $17.50.

8 As discussed below, many elements of a software program embodying a *de facto* standard will remain eligible for copyright protection, and the whole program may also receive protection. Our objection is to extending copyright to protect the elements of the program necessary to conform to or practice the standard.

...Copyright is the cost-effective approach only when the probability of [overprotection] is very small...

Economics of the Software Market

COSTS OF SOFTWARE DEVELOPMENT AND DISTRIBUTION

The production and distribution of computer software share several characteristics with other forms of intellectual property. First, most of the cost of developing a program (i.e., writing, testing, and debugging code) is independent of the number of copies distributed. The marketing of a software package (i.e., advertising and distribution) may enjoy significant scale economies as well. In contrast, the cost of duplicating and distributing the programs is negligible. As a result, once software is developed, the marginal cost of production is very low.

A second feature of software production is that much of the cost is sunk; a large fraction of development and marketing expenses cannot be recovered should the vendor decide to exit the business. The code usually has little value in other uses, and any learning acquired in the process can be applied only partially to other endeavors.

Third, software developers are not the only ones who make substantial investments in software products. Besides their out-of-pocket expenses for the software package itself and for the ancillary hardware and software to run it, users make considerable intangible investments. They acquire expertise while learning and operating the program, and create files and programs that are specific to the software package. These assets are rendered worthless if the vendor creates a new version of the software with specifications that are incompatible with the old version. Makers of compatible hardware components and software programs could find themselves in the same predicament. Expenditures by all these groups will be less sunk, in general, when industry technical standards ensure that their components are interoperable.

A final common feature is that the cost of developing a software package also depends on the stock of technologies that are technically and legally available to programmers. Developers of future generations of software benefit from the insights and the mistakes of current and earlier programmers. These benefits may derive from breakthroughs such as object-oriented programming that promise to improve the performance of all kinds of software, or advances in software that performs a narrow set of computing tasks. In either event, artificial restrictions on the use of past discoveries will necessarily raise the cost of current development. Restrictions on the use of general knowledge by subsequent innovators are particularly damaging to the social welfare and should be avoided.

NETWORK EXTERNALITIES

A full understanding of the software market must take into account the presence of network externalities. As noted above, network externalities occur when the value of a product or service to a buyer increases with the cumulative number of other buyers.[9] Each additional purchase raises the value to both current and future users.

The clearest example of a network externality is the telephone network. At one extreme, owning a phone has no value if you are the only person connected to the network. Telephone service becomes more valuable to each subscriber as more households are connected to the network. In the case of software there are several reasons why purchase of a software package delivers more value if many others use that same program. First, each user has more opportunities to share files and exchange expertise with other users. Second, a larger customer base can support production of a greater variety of complementary hardware and software by allowing timely recovery of the fixed costs associated with the development of these products. Since users benefit when hardware components and software applications used in conjunction with a software package become available, a person's willingness to invest in a hardware or software system will depend directly upon the cost and variety of complementary products.[10]

The user need not depend on one manufacturer to provide this variety, however. As long as products made by different vendors are compatible, users can mix and match components and create hybrid systems suited to their personal tastes.[11] For example, when evaluating whether to buy a particular home video game system, consumers consider the cost and variety of all compatible game cartridges, whether originally designed for that system or for some other system.

The important social value of compatibility among software programs is revealed in vendors' design strategies. They go to great lengths to ensure that upgrades are backward-compatible with earlier versions. More recently, software developers have marketed "suites" that contain an array of programs (e.g.,

9 The literature on network externalities has grown rapidly in recent years. Among the important articles are Katz and Shapiro [1985; 1992] and Farrell and Saloner [1985]. Gilbert [1992] provides a nice overview of the literature. Menell [1987; 1989] criticizes copyright doctrine for ignoring the importance of network externalities.

10 The courts have recognized the presence of network effects for personal computer operating system software; see *Apple Computer v. Microsoft Corp.*, 717 F.Supp. 1431.

11 The implications of this possibility are explored in Matutes and Regibeau [1988].

a word processor, a spreadsheet, and a database manager), and boast the ability to interchange files among the different applications. Presently, initiatives are underway to develop personal computer operating systems that will permit interchange of files among programs supplied by different vendors.

Economic Analysis of Copyright Protection

The economic conditions of the software market—massive scale economies in the form of large fixed sunk costs for development and marketing, plus extensive network externalities—have important implications for the efficient form and extent of intellectual property protection. Network externalities in particular lead us to recommend severely restricted use of copyright for software. First, when software programs achieve the status of a *de facto* standard, copyright protection allows the possibility of leveraging the monopoly into complementary hardware and software. Second, the same analysis applies to software interfaces used by popular programs. Finally, reverse engineering can realize efficiencies in incorporating advances into the next generation of software and provide a check on monopoly power that comes from first-mover advantages or from excessive intellectual property protection.

COPYRIGHT SHOULD NOT BE EXTENDED TO *DE FACTO* STANDARDS[12]

Creation of a *de facto* standard is a *joint* undertaking. It requires a first adopter, then a second, and a third, and so on. By itself, no hardware or software innovator can achieve market dominance by merely launching a product, no matter how early it arrives in the market (e.g., *VisiCalc*) or how large its sponsor may be (e.g., IBM's microchannel). A software package attains the status of *de facto* standard through the efforts of many sponsors: the original developer, the users who purchase the program, the makers of complementary hardware and software, and even the suppliers of compatible substitutes.

The multiplicity of sponsors of a *de facto* standard is a product of network externalities. Each additional purchase enhances the value of the program to current users and contributes to the standard's popularity by increasing its potential value for future consumers. A larger consumer base will also raise profit expectations of makers of compatible hardware and software products, whose expansion in turn further enhances the standard's use value.[13]

Even firms that offer competing products can contribute to the success of a standard. While these products may vie for the same customers as the dominant product, they also attract users who would otherwise opt for an alternative platform or who would not make any purchase at all. When users of the competitors' products consider upgrading, they become prime candidates for the purchase of the dominant product. By that time, they have made sunk investments in training and in complementary hardware and software that significantly lower the cost of switching to the dominant product.

As previously explained, all sponsors of a *de facto* standard—users of computer software and firms that supply compatible hardware and software products—make considerable investments in the *de facto* standard. They also bear the risks of its failure. The salvage value of these investments can be negligible if the prospective standard should fail to achieve critical mass, or if its specifications change along the way.[14]

For efficiency's sake, each adopter should receive a return equal to the incremental benefits that its purchase confers on others over time. The nth purchaser should have claim to the incremental benefits that its purchase provides for the current $n-1$ users through the remainder of the product's useful life. A purchaser should also have claim to the stream of incremental benefits enjoyed by the $n+1$st purchaser, the $n+2$nd purchaser, and so on.[15] Furthermore, it is possible that, by forgoing a purchase, a user raises the chance that other users will abandon the product, possibly switching to a more promising one.[16]

In principle, a *de facto* standard's original developer may find it profitable to "internalize" the incremental benefits of the many sponsors through the use of "penetration pricing." By setting lower prices to initial adopters, the innovator compensates them for the benefits they provide later adopters and for the risk they bear that the standard will not achieve critical mass, and even for the risk that the original developer will eventually exploit their locked-in position. Often, software vendors vary prices among customers and over time partly to reflect costs of serving them as well as their willingness to pay. Vendors are also known to adjust licensing terms over time to suppliers of complementary hardware and software.[17]

In practice, however, a firm granted ownership over a *de facto* standard may find it profitable to set prices well above marginal cost, which will ineffi-

12 Here we are talking about the standardized elements of the program necessary to practice, or conform to, the standard. This does not include those elements of the program that are not essential to practice the standard, nor do we mean to imply that the program as a whole is unprotectable.

13 While a purchase benefits current and future users, it may hurt users of a competing incompatible standard which, in the limit, could be abandoned, stranding all of its adopters.

14 In response, these firms may take costly measures to reduce the risk associated with adhering to the monopoly-sponsored standard; e.g., they may design their products (at an additional cost) to ensure they can be "ported" to different hardware or operating system platforms.

15 This conceptual experiment envisions purchasers who exclusively occupy a position in the sequence of purchases. Should a user forego a purchase, and if another user automatically enters and makes the purchase instead, then the incremental benefit of a particular user's purchase is significantly reduced. This would be the case, for instance, if there was an unlimited number of possible consumers.

16 While, in theory, these direct benefits could be measured, it could be particularly difficult, for example, to estimate benefits from heightened expectations that the prospective standard will exceed some critical mass.

17 For example, home video game manufacturers are reported to introduce new hardware platforms with especially attractive licensing terms to game developers; see Sheff [1994].

...assignment of copyright can greatly overcompensate the original creator of a de facto standard.

ciently discourage use. To make matters worse, the monopoly sponsor will be tempted to charge initial customers higher prices to extract a greater portion of their higher willingness to pay. The tendency to "price skim" (i.e., price discriminate over time) leads a monopolist to set prices that fall, not rise, over time.[18] Lastly, information costs impose severe limits on the degree to which the monopolist can price discriminate among buyers based on their incremental contribution to network benefits.

Given the price mechanism's limited ability to support efficient adoption incentives when network externalities are present, assignment of copyright can greatly overcompensate the original creator of a *de facto* standard. Moreover, monopoly rights may be given to the innovator of what is later revealed to be an inferior technology. This happens with surprising frequency, even without government intervention,[19] and copyright protection for *de facto* standards will raise that likelihood.

INTERFACE SPECIFICATIONS SHOULD NOT BE COPYRIGHTABLE

Misassignment of copyright can be especially pernicious in the case of interface specifications. To understand why, it is necessary to refer to the copyright doctrine of merger. The "merger doctrine" holds that when idea and expression are one and the same, or when there are only a small number of expressions of an idea, the expression is not protectable under copyright, as control over the expression would be tantamount to control over the idea. In economic terms, such a copyright would give unwarranted monopoly power over the idea. Moreover, control of the initial idea may prevent, or limit, the ability of others to build on it, allowing an extension of the monopoly to other complementary products or to later improvements on the idea.

It is widely agreed that this basic merger doctrine of copyright should and does apply to software. In a network context, the same analysis implies that, if a program achieves the status of a *de facto* standard, then the courts should not rule out a merger of idea and expression simply because other, now inferior, expressions of the same idea were, and still are, available. To see this, consider the following not-so-hypothetical example.

Suppose that *ex ante* there are numerous ways to efficiently compose a piece of software, such as a spreadsheet, and that several companies develop a spreadsheet program. Consumers have a very strong preference for standardization of software offerings, so that any spreadsheet package that obtains a lead in installed base quickly becomes the uniform selection of subsequent consumers. Then, *ex post,* any innovative and improved version of a spreadsheet must be compatible with the established standard.[20]

If the copyright covers the interface specifications with which the spreadsheet interacts with other software, hardware, and with users, then the copyright on the initial standard essentially forecloses competition for future spreadsheet generations. No social purpose is achieved by this grant of a valuable copyright monopoly.[21] Any of the spreadsheet configurations would have been equally valuable as the standard, since *ex ante* there were numerous ways to express the idea(s) embodied in the spreadsheet. In that case the creation of a copyright monopoly that covers interfaces is not in society's interest.[22]

More generally, since copyright should provide protection only in circumstances where there is a low probability that an unwarranted monopoly will have been awarded (i.e., a small chance of a "false positive") and since control over interfaces can easily convey market power and control over substitutes and complements, interface specifications should not receive the automatic and uncritical protection afforded by copyright.[23]

REVERSE ENGINEERING SHOULD BE ALLOWED

Much of the recent controversy concerning software copyrights has involved reverse engineering. Various methods of reverse engineering (including so-called black box techniques, disassembly, and decompilation) are apparently widely used for a variety of purposes by customers, designers of complementary

18 Cabral et al. [1994] show that in most cases a monopoly provider of a product that confers network externalities will never set prices that increase over time.

19 See Arthur [1989] and David [1985].

20 *Ex ante* here refers to the period before sunk costs are incurred. *Ex post* refers to the period after sunk costs have been incurred.

21 The analogy is with the monopoly power that economies of scale can confer upon pipelines or electric utilities. The usual policy response to massive economies of scale in production has been utility regulation to prevent the exercise of monopoly power. In computer software, where network effects correspond to economies of scale on the demand side, limits on the extent of copyright protection are all that is necessary to prevent the exercise of monopoly power.

22 Menell [1987; 1989] shares our concern that copyright protection should not extend to interfaces, at least for software achieving the status of a *de facto* standard. Unlike us, he has reservations as to whether the correct result can be reached applying copyright's merger doctrine, and he suggests alternative means such as compulsory licensing, applying the fair use doctrine to limit the protection obtainable for interfaces, or creating a new category of intellectual property for software.

23 Copyright protection is automatic in the sense that its grant occurs at the request of the developer. Copyright protection is uncritical in that application is not reviewed to establish novelty or nonobviousness as in the case of patents.

applications, and direct competitors.[24] Reverse engineering to develop directly competing products is the most controversial use. There are several reasons why reverse engineering should be allowed, even by suppliers of competing products.

First, traditional copyright analysis implies that reverse engineering to achieve interoperability should be lawful in markets with network externalities. To continue with the previous hypothetical example, now suppose that the interface standards are not covered by copyright, but that the developer of the initial spreadsheet chooses not to make all interfaces public. We retain the assumption that users invest heavily in their spreadsheet applications, and would be unwilling to switch to another, nonstandard, spreadsheet. Companies seeking to compete for future generations or improved versions of the spreadsheet must therefore deduce the interfaces and internal operations of the dominant spreadsheet (or license the software at monopoly rates).

Arguably the most cost-effective means of deduction involves making a "copy" (via decompilation or disassembly) of the standard spreadsheet. Should would-be competitors be prevented by copyright from doing so? No. *Ex post,* it is consumer behavior and not the inherent superiority of the initial spreadsheet that has created a merger of expression and idea since, as a practical matter, second generation spreadsheets must be fully compatible with the first generation standard.

Copyright law should allow such reverse engineering either on the grounds of an *ex post* merger of expression and idea, or on the grounds that these methods of reverse engineering merely allow competitors to (perhaps only partially) restore the *ex ante* circumstances wherein there is nothing special about the first spreadsheet.[25] The standard's creator cannot have it both ways; that is, argue that the copyright is valid because *ex ante* there were many viable ways of expressing the idea, and then use copyright to prevent competitors faced with an *ex post* merger of idea and expression from seeking to restore via reverse engineering the circumstances that made the copyright valid.[26]

A second reason to allow reverse engineering is that markets may be biased against open systems.[27] This is because in network markets proprietary systems may have a strategic advantage unrelated to efficiency. Proprietary systems can more easily engage in penetration pricing that can open systems, whose many sponsors must somehow agree who will bear what portion of early sacrifices in profits, and who are unable to recoup losses through higher prices in the face of an onslaught of new entrants. If markets are biased against open systems, that bias should not be exacerbated by software copyright principles that inhibit the evolution of open systems out of closed ones.

Finally, reverse engineering is necessary to ensure that copyright should not extend to *de facto* standards. Otherwise firms "owning" the *de facto* standard will tend to release insufficient information to provide compatibility. For example, Microsoft regularly conceals interface standards, as was revealed in the recent Stac litigation.[28] Absent an ability to judge for themselves (via reverse engineering) whether the release of interface information was adequate, competitors would be at the mercy of the standard owner, whose incentive is often not to release (at least not in a full or timely fashion) interface information that others need in order to compete.[29]

Software Copyright Case Law

In this section we apply our economic analysis of software copyright to several recent cases. From an economic perspective, the *Altai* case seems correctly decided and even uncontroversial. In contrast, the logic of the *Paperback* court in finding liability appears at odds with our analysis. Fortunately, the recent *Borland* decision restored valid economic logic to disputes involving alleged infringement by a direct competitor in markets where user investments and network externalities are important.

The *Sega* and *Nintendo* cases, in our view, involve quite different issues that have not been recognized. While network externalities may be somewhat relevant in those cases, we suspect that the primary reason Sega and Nintendo attempted to restrict the sale of competing compatible games was not to deny network economies of scale to potential competitors. Rather, their most likely motive was price discrimination. We argue that using copyright to maintain price discrimination is inappropriate and probably unwise. Thus, the *Sega* and *Nintendo* courts arrived at the correct conclusion, albeit perhaps for the wrong reasons.

24 At one extreme is literal duplication of the object code. At the other is decompilation or disassembly for the purpose of academic research. Often entrants may reverse engineer programs in order to assure compatibility with complementary applications or for future compatibility with successive versions of the program. They may seek access to noncopyrighted portions of the program. Users may also disassemble a program so as to customize its operation to conform to their special needs or operating environment.

25 Of course, competitors should not be allowed to sell copies of the expression (code) of the initial spreadsheet standard. Competitors would not have to do so to compete on a level playing field with the incumbent, since by assumption *ex ante* there are equally acceptable ways of expressing the idea.

26 By similar argument, a software copyright holder should not be allowed to prevent competitors from utilizing syntax and command structures that users have learned from the initial spreadsheet. It would, however, be inefficient, and therefore should be illegal under trademark or copyright laws, to allow copying of screen displays. Copying of displays is not necessary to achieve compatibility, nor to allow users to continue employing the commands they have mastered, but it may create confusion as to who developed the software (and is responsible for supporting it).

27 See Katz and Shapiro [1992] and Farrell [1989].

28 In that litigation, Microsoft argued that undocumented calls in Windows were its own trade secrets. This position was inconsistent with its prior announcements to competing developers of Windows applications that Microsoft's in-house applications operations were not favored *vis-a-vis* independent software houses.

29 Menell [1987] proposed compulsory licensing of software programs that have emerged as industry standards. Such a policy probably could not substitute for the right to reverse engineer to achieve compatibility, given the standardholder's incentive for less-than-complete disclosure of relevant information.

Unlike many commentators, however, we do not see that the case for allowing disassembly in order to make complementary products is more compelling than the case for allowing the practice in order to develop directly competing products. Indeed, in markets where network effects are important, the latter case is even stronger than the former.

COMPUTER ASSOCIATES V. ALTAI AND LOTUS

Several recent cases have addressed the extent to which owners of established software products can invoke copyright infringement against new direct (i.e., horizontal) competitors. There is no disagreement that exact copying of object code constitutes infringement. These cases have all involved the extent to which new competitors can use a logical structure or user interfaces that are identical or similar to those of the incumbent.

The reasoning underlying the *Altai* decision is sound.[30] *Altai,* however, did not raise the issues of user investment or network externalities. Network effects were apparently not important in the market so the distinction between *ex ante* and *ex post* merger of idea and expression did not arise. The court decided that given the tight constraints imposed by the operating environment and the utilitarian task to be accomplished, there was an *ex ante* merger of idea and expression.[31] Earlier we defended this established principle of copyright law on economic grounds.

Several critics agreed with Computer Associates' contention that there was no *ex ante* merger of idea and expression. They reasoned that Altai was free to develop and market its product in a non-IBM environment, to write separate applications for each IBM operating system, or to develop from scratch its own operating system for IBM mainframe computers. In practice, however, these alternatives would block Altai's access to a large number of customers. Not only would this harm those customers, but the high fixed costs of software development imply that denial of access to a significant share of the market would increase the average cost of supplying the remaining customers, perhaps so much that they could not be served economically, and entry would be effectively blocked altogether.

Two important cases involving Lotus have addressed whether the commands and the logic of the command structure in a software package should receive copyright protection. In *Lotus v. Paperback,* the defendant sold a spreadsheet program with a user interface that was very similar to the Lotus interface. The court found that although several individual elements of the interface could not be copyrighted because they were merged with the idea of the spreadsheet, the command structure of the interface was protectable because numerous other ways to set up the spreadsheet's user interface were feasible. The defendant argued that, given Lotus' dominant position as the spreadsheet standard, customers would not regard interfaces that required user retraining as reasonable alternatives. The court rejected the argument, saying that it turned copyright principle on its head and that Lotus should not be penalized for its success.

It would appear that the court implicitly assumed that 1-2-3's success did not increase the real cost to Paperback of expressing the unprotected idea of spreadsheets, i.e., that 1-2-3 became the standard solely or primarily because of its own efforts. We would agree that, so long as the presence of a copyright-protected 1-2-3 reduced the price that consumers were willing to pay for Paperback's spreadsheet simply because 1-2-3 offered a superior product,[32] Paperback would have no case. But if Lotus imposed a real cost on Paperback by preventing potential customers of Paperback from taking advantage of network economies of scale (benefiting users of both 1-2-3 and the Paperback spreadsheet), then Lotus' attempt to reduce compatibility through its exercise of copyright protection harmed both competitors and consumers.

In those circumstances copyright protection that limited rivals' ability to achieve 1-2-3 compatibility would raise the cost they would face in developing and marketing competing spreadsheets, even those with significant new and enhanced features.[33] Unless the court were certain that no network externality existed (did it even consider the possibility?), copyright protection of this sort is unjustified (though patent protection might be justified).[34]

Our problem with Judge Keeton's logic in *Paperback* can also be seen by noting that he analyzed liability at the wrong time. Consistency requires that

30 *Computer Associates International, Inc. v. Altai, Inc.*, United States Court of Appeals for the Second Circuit, January 9, 1992.

31 Computer Associates and, later, Altai sold job-scheduling programs that operated within IBM mainframe environments and contained program modules that allowed the same applications software to perform on different operating systems in that environment. The alleged infringement involved the converter or adaption routines that allowed the applications software to run on the different operating systems.

32 In economics jargon, in these circumstances the presence of a very successful 1-2-3 would impose a negative "pecuniary" externality on the makers of competing software. Economists generally would recognize that no market failure is indicated, and no correction is called for, if Lotus has gained by making its product more attractive (lower prices, higher quality) to consumers, even though this may reduce the amount that those consumers would be willing to pay for rival products. Imposing pecuniary externalities on one's rivals is just part of the normal process of competition.

33 Again in economics jargon, such copyright protection would impose a "real" or "technological" externality on rivals. A real negative externality (pollution is the standard example) imposes net costs on the rest of society, as would be the case to the extent that Lotus has gained not by making its product more attractive to consumers but by making its rival's products *less* attractive. Real externalities are a source of market failure, and correction is called for unless the social costs of the correction exceed the costs of the externality. Here there are no social costs of correcting for the eternality; indeed, the social costs are actually negative, since correcting the externality simply involves not allowing companies to use the socially-provided justice system to enforce copyright principles that could impose substantial real externalities.

34 Lotus claimed that a substantial fraction of its development costs for *1-2-3* were incurred in designing the user interface and that its research made significant advances in the state of the art for user interfaces. If true, Lotus may have been entitled to patent protection. For an insightful discussion of alternative legal regimes that could be desirable to induce the proper level of investment in software if existing legal regimes are inadequate, see Samuelson et al. [1994] and Samuelson [1995].

liability and harm be assessed at the same time.[35] Judge Keeton argued that the copyright survived challenge under the merger doctrine based on the fact that alternative ways existed for Paperback to express the ideas in its spreadsheet. Therefore, Paperback was guilty of infringement. However, Lotus' argument that it had been damaged was based not on the *ex ante* situation before users had invested heavily in learning the Lotus spreadsheet, but on the *ex post* situation where, precisely because of heavy user investment, compatibility with Lotus was very important for competing spreadsheets.[36] *Ex post,* there had been a significant merger of expression and idea, and this merger should have been recognized by the court in judging the extent to which Paperback had available realistic alternatives in each of the areas where Lotus claimed Paperback's spreadsheet was too similar to its own.

The *Lotus v. Borland* litigation addressed the proper extent of copyright protection over "proprietary" commands and command structures. Borland included the capability to display, execute, and edit Lotus 1-2-3 macros in its Quattro Pro spreadsheet. Lotus claimed that Borland infringed its copyright over the menu command hierarchy which was copied into the Quattro Pro program. Borland responded by arguing that the 1-2-3 menu structure is a "system, method of operation, process or procedure" for controlling a computer program, and hence is uncopyrightable under Section 102(b) of the Copyright Act.

Once again, Judge Keeton heard Lotus' complaint.[37] He ruled that 1-2-3's menus and submenus arranged in a specific order constituted a protectable expression of the underlying idea of controlling the spreadsheet program. He arrived at this conclusion by separating the abstract notion of a method of operation (i.e., software control of a spreadsheet program) from its copyrightable implementation (i.e., 1-2-3's menu structure).

In its review of the district court's decision, the appeals court concluded that such a separation was unwarranted.[38] Instead, it held that the menu structure was a program interface that was physically separable from the spreadsheet program but essential to its operation. This reasoning led it to agree with Borland that the 1-2-3 menu structure was a "method of operation," and so was unprotectable.

It is instructive that, while neither decision paid attention to the presence of network externalities, the appeals court did recognize the relevance and importance of sunk investments by users in learning the interface and in writing specialized complementary software. Customers spend considerable sums (by some estimates seven times Lotus' investment in 1-2-3) to create customized macro programs, and would incur significant switching costs if those programs had to be rewritten for another spreadsheet program. The court noted that:

> Under the district court's holding, if the user wrote a macro to shorten the time needed to perform a certain operation in Lotus 1-2-3, the user would be unable to use that macro to shorten the time needed to perform that same operation in another program. Rather, the user would have to rewrite his or her macro using that other program's menu command hierarchy. This is despite the fact that the macro is clearly the user's own work product.[39]

Judge Boudin's concurring opinion, in particular, also pointed out that the 1-2-3 interface had become the *de facto* standard largely or even primarily because of continued investments over time by users:

> A new menu may be a creative work, but over time its importance may come to reside more in the investment that has been made by users in learning the menu and in building their own mini-programs—macros—in reliance upon the menu.[40]

Judge Boudin went on to question why Lotus should be allowed to use the advantage it gained from 1-2-3's status as the *de facto* standard to seize users' sunk investments:

> So long as Lotus is the superior spreadsheet—either in quality or in price—there may be nothing wrong with this advantage. But if a better spreadsheet comes along, it is hard to see why customers who have learned the Lotus menu and devised macros for it should remain captives of Lotus because of an investment in learning made by the users and not by Lotus.[41]

Critics of this argument may contend that even if users appear to have undertaken substantial sunk investments in learning the program and in writing complementary software, some or all of these costs may have been borne by the developer. This might occur even if the developer did not directly compensate users for such investments. As noted above

35 See Fisher and Romaine [1990]. This influential article poses the question: what amount of damages is due to the owner of a high school yearbook containing Janis Joplin's autograph if the yearbook is stolen? The answer depends on *when* the yearbook was stolen. If it was stolen before Joplin became famous, its market value was small. If it was stolen after Joplin became famous, its market value was high. Fisher and Romaine argue persuasively that it is very inefficient to allow plaintiffs to sue for windfall damages, as would be the case if the yearbook was stolen before Joplin became famous, but damages were claimed on the value of the yearbook after her rise to fame. The plaintiff would then be made more than whole, since the loss at the time of the crime was far smaller. Moreover, a rule allowing claims for windfall damages creates perverse incentives for litigation. No efficiency purpose is served by the litigation unless the expected harm from the wrong at least exceeds the plaintiff's costs of litigation at the time the wrong occurred.

36 See Gandal [1994] for empirical evidence as to the importance of Lotus compatibility in the spreadsheet business.

37 Lotus Development Corp. v. Borland International, Inc. 831 F. Supp. 223 (D. mass 1993).

38 Lotus Development Corp. v. Borland International, Inc., U.S. Court of Appeals for the First Circuit, March 1993.

39 Id. at 27-28.

40 Id. at 32.

41 Id. at 37.

in our discussion of penetration pricing, a *de facto* standard's original developer may find it profitable to "internalize" the incremental benefits of the many sponsors by setting prices to initial adopters that are lower than those that would have maximized its profits in each period. Indeed, if users expect that the initial developer will eventually appropriate the value of those investments through its control over a *de facto* standard, users would rationally require compensation *ex ante* in the form of lower—perhaps even negative—prices for the initial software. With perfect and costless information and foresight, users could thus prevent the developer of a *de facto* standard from opportunistically seizing the value of their investment.

Given that information available to users is neither costless nor perfect, however, the ability of users to force developers to fully compensate users for their sunk investments in complementary products through penetration pricing may be too limited to be much relied on in practice. In addition, as discussed above, for products with many users with differing levels of willingness to pay (i.e., different reservation prices for the product) setting prices that are initially very high and then decline (i.e., price skimming) can be the most profitable strategy for the developer. Moreover, in such an imperfect world new users (those who have not yet chosen a spreadsheet program) as well as old users may be harmed if entrants cannot compete on an equal basis for sales to locked-in users. Given the high fixed costs involved in software development, denial of access to a significant share of the market will increase the average cost to an entrant of supplying just the new customers, allowing the incumbent to raise prices to new customers as well as to old.

Thus, while the presence of sunk user investments and scale economies may not be sufficient to ensure that extending copyright protection for software to cover *de facto* standards or interface specifications would be anticompetitive in every instance, the likelihood of such an outcome is sufficiently high that the automatic and uncritical protection provided by copyright is inappropriate when users have made significant sunk investments.

The presence of sunk user investments is not, however, the only valid reason for denying copyright protection to interface specifications and *de facto* standards. As discussed above, the case is made even stronger when network externalities are present. In that event, broad copyright protection can allow the developer to appropriate the value of network externalities enjoyed by current customers. If, in addition, copyright protection raises the cost to prior users of switching to a competing product and allows the first mover to block access to a sufficiently large number of customers that an entrant would be unable to cover its fixed costs, then such protection also grants to the first mover *ex ante* monopoly power over new users in the future.

Furthermore, the larger the first mover's installed base of users, the greater compensation both new and old customers will demand to switch to the new product whether or not they have sunk investments in the current interface specification or *de facto* standard. Where the installed base of users is large, coordinating the simultaneous change-over of a significant share of those users to a new *de facto* standard may in practice be an impossible task for an entrant. This implies that the appeals court's rejection of copyright protection when it allows appropriation of user sunk investments should be extended to rejection of copyright protection when network externalities are present, even absent any sunk investments by users.

Interestingly, in the case of Lotus 1-2-3, both user investments and network externalities were sufficient to reduce many *ex ante* equivalent means of expression to a single means of expression that is *ex post* uniquely superior to all others. This suggests that an appropriate and consistent economic basis for denying copyright protection to 1-2-3's menu structure can be found in the legal doctrine of merger. As we have stated earlier:[42]

> We would interpret Borland's position as a statement that network externalities have transformed an *ex ante* large number of alternative expressions into an *ex post* merger of the idea with one expression, and that much of the sunk investment incurred to create that advantage for the Lotus product was in fact incurred by consumers. Hopefully, the Borland court will recognize the importance of these arguments, and grant them greater deference than did the Paperback court. Aspects of 1-2-3 have become *de facto* standards for spreadsheets, and the question the court should attempt to answer is whether the alleged infringement involves a *de facto* standard. If so, copyright restrictions preventing others from conforming to the standard are inappropriate.

While the term "merger" never appears in the decision, the Court's statements carry at least the flavor of the merger doctrine.[43] While it is not "idea" and "expression" that is merged here, the cause of the merger is the same: given network externalities and users' sunk investments in the implementation, the cost of introducing an alternative implementation (i.e., the users' switching cost) is increased significantly, leaving only one economical implementation of the uncopyrightable material.[44]

---

42 Warren-Boulton, Baseman and Woroch, [1995, 23–24].

43 Expression is not copyrightable because it is *part of* Lotus 1-2-3's "method of operation." Id., 23; "Thus the Lotus command terms are not equivalent to the labels on the VCR's buttons, but are instead *equivalent to* the buttons themselves." Id., 27. Emphasis added.

44 Similarly, the *scenes a faire* doctrine prevents the first mover from appropriating the value of investments made by others (e.g., when subsequent authors continue the development or recognition of a stock character), or the value to new users (readers) created by its use by other users (recognition of a stock character), which we would regard as a network externality.

RECENT CASE LAW: THE NINTENDO AND SEGA CASES

The *Nintendo* and *Sega* cases established a "fair use" copyright exception for developing complementary products. Defendants produced competing games that could run on Sega's or Nintendo's game hardware, and succeeded in making their games compatible with the Sega or Nintendo hardware by using disassembly to crack the "lock-out" code in the software embedded in the game hardware. Both courts determined that this reverse engineering was lawful to the extent that other means of attaining compatibility (or learning the unprotected ideas in the embedded software) were not available. These decisions have been criticized on the grounds that Accolade and Atari were competitors of Sega and Nintendo at the game level: Accolade and Atari were not providing an entirely new use for the hardware, in which case the hardware manufacturers might have reacted differently.

If a company develops a new (or even simply an improved) compatible and complementary product, it may be hard to see how the copyright holder can claim to have been harmed, since the demand for the initial product will increase because of this new product. But the economics of vertical control (as control over a complement is usually described) is quite complicated. There are a variety of reasons why an upstream company would seek to control a downstream market.

A common motivation, which we suspect is the primary explanation for the behavior of Sega and Nintendo, is price discrimination. Since the potential buyers of a game system are likely to vary considerably in terms of the maximum amount that they would be willing to pay for a game system, a game manufacturer would like to price discriminate, charging users who place a high value on the game system a higher price than users with lower valuations. A game manufacturer cannot do this simply by charging different customers different prices for the hardware. The manufacturer can neither easily identify which customers would be willing to pay a high price and which a low price, nor can it prevent those receiving it at the low price from reselling it to those who would pay the higher price.

The standard solution to this problem is to find a complementary good, the demand for which can serve as a measure of how much the customer values the system. The manufacturer then requires that the customer buy the complement from the manufacturer, and marks up the price of the complement. For example, an inventor of a new type of razor might reasonably expect that customers who use many blades would be willing to pay more for the razor than customers who use very few blades. The inventor who can ensure that no one else can supply blades that are compatible with the new razor can combine a low price for the razor with high profit margins on the blades. Much higher overall profits will thus be realized by extracting higher net revenues per razor from high-intensity users without having to forego profitable sales to low-intensity users. Similarly, for game systems, a customer's demand for games is a reasonable indicator of his or her reservation price for the game system. A game manufacturer who can prevent others from providing compatible games (or charge suppliers of compatible games a high license fee) will find it profitable to take profits at the game level rather than at the hardware level.

The critical problem, of course, is excluding others from supplying the complementary product, especially given the high margins set by the manufacturer for the games. When a manufacturer attempts to exclude others by contract with buyers, this is referred to as a tying arrangement, and often leads to dire antitrust consequences. Here, Sega and Nintendo appear to have attempted to use the copyright laws to exclude other suppliers of compatible games in order to price discriminate among users.

While economists regard the welfare effects of such price discrimination as indeterminate *a priori*, and would thus generally support a rule of reason approach to this kind of action, antitrust case law has taken a much harsher view.[45] The welfare effects of vertical control are ambiguous. Some motives for vertical control generally result in its being beneficial (maintaining the quality of complements or preventing free-riding). Other motives have ambiguous effects (price discrimination), and still others generally result in bad or inefficient effects (raising rivals' costs or increasing entry barriers). Making the correct diagnosis is often very difficult. Whatever the determination, however, it would be ironic indeed for a supplier to achieve through copyright something that would be illegal if achieved through agreement.

There are three possible approaches to this complex issue, each corresponding to an established antitrust rule:

(1) *Per se* illegality for decompilation and disassembly (or for unauthorized copying in general), which corresponds to a rule of *per se* legality for vertical control.

(2) Some "rule of reason" inquiry into the economic effects of banning or limiting disassembly, corresponding to a rule of reason for vertical control.

(3) *Per se* legality for decompilation and disassembly, corresponding to a rule that vertical control must be achieved through other means.

Given the *a priori* uncertainty about the effects of copyright rules that allow control over complements, the first option could only be defended if it were clear that disassembly or decompilation has undesirable

45 The antitrust cases involved contracts (a contractual tie) that compelled users to purchase the complements from the manufacture. Sega and Nintendo sought to realize a similar result through creating a technological tie. Indeed, they apparently had to incur extra cost to create the technological tie. See Warren-Boulton [1978] for a discussion of the issue generally, and Greenstein [1990] for a discussion of the issue in the context of software.

effects in almost all circumstances and that it is too costly to determine when the effects might be benign. But given that end-users often engage in these practices to make software more useful to them—hardly a practice against the interest either of the software developer or of society—this position appears untenable.

Either a "rule of reason" or *per se* legality for decompilation and disassembly can be defended in principle, although a rule of reason could impose substantial uncertainties as well as legal and enforcement costs. Absent some estimate of the costs and benefits from a rule of reason approach, therefore, *per se* legality for decompilation and disassembly appears to be the most appropriate rule.[46]

Conclusion

We have argued that copyright protection should be provided only when the probability of a false positive (i.e., granting an unwarranted monopoly) is *de minimis*. But, as we have seen, when network externalities or significant user investment are present the control over substitutes made possible by copyright of interface specifications or by *de facto* standards can significantly harm competition and reduce welfare. In addition, when copyright is used to control complements, the welfare effects of such vertical control are at best ambiguous. We conclude, therefore, that copyright should not be used to block compatibility with a rival's product, whether that product is a complement or a substitute, and that copyright protection should not be extended to interface specifications or to *de facto* standards. Fortunately, with some minor exceptions, recent case law has taken steps in this direction by helping to clarify legal principles that prevent producers from gaining excessive protection under copyright.

As noted in our introduction, we do not regard the specific policy decisions evaluated in this article—e.g., copyright of interface specifications or *de facto* standards and decompilation—as reflecting a battle between creative producers and later arrivals attempting to free-ride on their work. All the participants in this debate and in its associated litigation create value. The debate is between producers who came first and producers seeking to both build on and advance the past work of the others. For the latter group, setting appropriate standards that neither overprotect nor underprotect software is important and desirable, for they can expect their role, and therefore their self-interest, to fluctuate over time between that of pioneer and that of developer. Our goal here has been to contribute to establishing a legal framework of intellectual property rights that neither robs innovators of their contribution nor handicaps those who come after them. Creating a level playing field for all producers is both equitable and critical to the continued expansion of a vital and intensely competitive industry. **SV**

ACKNOWLEDGEMENTS:

We are grateful to Peter Choy, Russ Wayman, and Christine Owens for their comments on an earlier draft of this paper. We also thank John Burton, Yale Braunstein, Pamela Samuelson, and Paul David for useful discussions.

References

ARROW. K., 1962. Economic welfare and the allocation of resources for invention. In *The Rate and Direction of Economic Activity*, Princeton University Press. Princeton, N.J.

ARTHUR, B., 1989. Competing technologies, increasing returns, and lock in by historical events. *Econ. J.* (March).

BARZEL. Y. 1968. Optimal timing of innovation. *Rev. Econ. Stat.*

CABRAL, L. SALANT, D., AND WOROCH. G. 1994. Monopoly pricing with network externalities. Draft, Univ. of California, Berkeley.

DAVID, P. 1985. Clio and the economies of QWERTY. *Am. Econ. Rev. 75*, 1 (May), 332–337.

FARRELL, J. 1989. Standardization and intellectual property. *Jurimetrics J.* (Fall), 35–50.

FARRELL, J. AND SALONER, G. 1985. Standardization, compatibility, and innovation. *Rand J. Econ. 16*, 1 (Spring) 70–83.

FARRELL, J. AND SALONER, G. 1992. Converters, compatibility, and the control of interfaces. *J. Indust. Econ. XL*, 1 (March) 9–34.

FISHER, F. AND ROMAINE R. C. 1990. Janis Joplin's yearbook and the theory of damages. *J. Account. Auditing Finance 5*, 1 (Winter) 145–157.

GANDAL, N. 1994. Hedonic price indexes and an empirical test for network externalities. *Rand J. Econ.* (Spring).

GILBERT, R. 1992. Symposium on compatibility: Incentives and market structure. *J. Indust. Econ. XL*, 1 (March) 1–8.

GREENSTEIN, S. 1990. Creating economic advantage by setting compatibility standards: Can physical tie-ins' extend monopoly power? *Econ. Innovation New Technol. 1*, 63–83.

KATZ, M. AND SHAPIRO, C. 1985. Network externalities, competition, and compatibility. *Am. Econ. Rev. 75*, 3 (June), 424–440.

KATZ, M. AND SHAPIRO, C. 1992. Product introduction with network externalities. *J. Indust. Econ. XL*, 1 (March), 55–83.

LANDES, W. AND POSNER, R. 1989. An economic analysis of copyright law. *J. Legal Stud. XVIII*, (June), 325–363.

MATUTES, C. AND REGIBEAU, P. 1988. Mix and match: Product compatibility without network externalities. *Rand J. Econ. 19*, 2, (Summer), 221–234.

MENELL, P. 1987. Tailoring legal protection for computer software. *Stanford Law Rev. 39* (July), 1329–1373.

MENELL, P. 1989. An analysis of the scope of copyright protection for application programs. *Stanford Law Rev. 41* (May), 1045–1104.

MILLER, A. 1993. Copyright protection for computer programs, databases, and computer-generated works: Is anything new since CONTU? *Harvard Law Rev. 106*, 5 (March), 977–1073.

SAMUELSON, P. ET AL. 1994. A manifesto concerning the legal protection of computer programs. *Columbia Law Rev. 94*.

SAMUELSON, P. 1995. An entirely new regime is needed. *Comput. Lauyer 12* (Feb.), 11–17.

WARREN-BOULTON, F., BASEMAN, K. AND WOROCH, G. 1995. Copyright protection of software can make economic sense. *Comput. Lauyer 12*, 2 (Feb.), 10, 18–28.

WARREN-BOULTON, F. 1978. *Vertical Control of Markets: Business and Labor Practices*, Ballinger.

46 Under this proposal, a firm could still attempt to exercise vertical control, but it would have to choose a more transparent method.

Frederick Warren-Boulton
Vertical Control of Markets (1978)

costs of using the market, or that government regulation or standards of legal liability prevent bargains from being made between independent parties that would lead to an efficient solution.[11]

For many cases of ownership externalities, less complete forms of vertical control than integration may be sufficient. For example, where the manner of use of a product by purchasers may affect the reputation and future sales of the producer, the producer may want users to uphold certain standards in the use or maintenance of the product. While the producer could simply reward adherence to such standards, the information and control costs of doing so may be high. Complete control through vertical integration over all aspects of the purchasing firm may also be expensive. The producer may therefore choose simply to tie the sale of the product to a maintenance contract,[12] either explicitly or by renting the product rather than making an outright sale; or to tie the product to sales of "approved" parts;[13] or to operate a franchise agreement. Similarly, if promotional efforts are best made at downstream levels, and if direct payment for such activities is difficult to calculate, then resale price maintenance or an exclusive franchise arrangement may be used to ensure that the return to such promotional activities accrues to the promoter.

Technical Externalities

If the problem is technical externalities, so the production of the input takes place at declining average cost, use of the market involves two problems. First, the input must be transferred at its marginal cost. Second, continued existence of the firm must be possible under marginal cost pricing. These problems would appear to be solved by integration between the (single) input producer and the users of the input. The new firm will set the internal transfer price of the input equal to its marginal cost, and the input-producing subsidiary can continue to exist despite accounting losses.

Integration, however, simply shifts the problem to a different level if the final output is now produced at falling average costs. For integration to provide a solution requires that enough other inputs are supplied internally at rising average cost so that the average-cost curve for the final product is not falling over the relevant output range. In effect, the accounting loss from production of the declining average-cost input is made up by rents from the increasing average-cost inputs. If inputs are used in fixed proportions, this solution can be shown graphically. Suppose that a final good, X, is produced competitively using two inputs, A and B, in fixed proportions. For simplicity assume that combining one unit of A with one unit of B results in one unit of X. In Figure 2-1, AC_a, AC_b, and AC_x are the average-

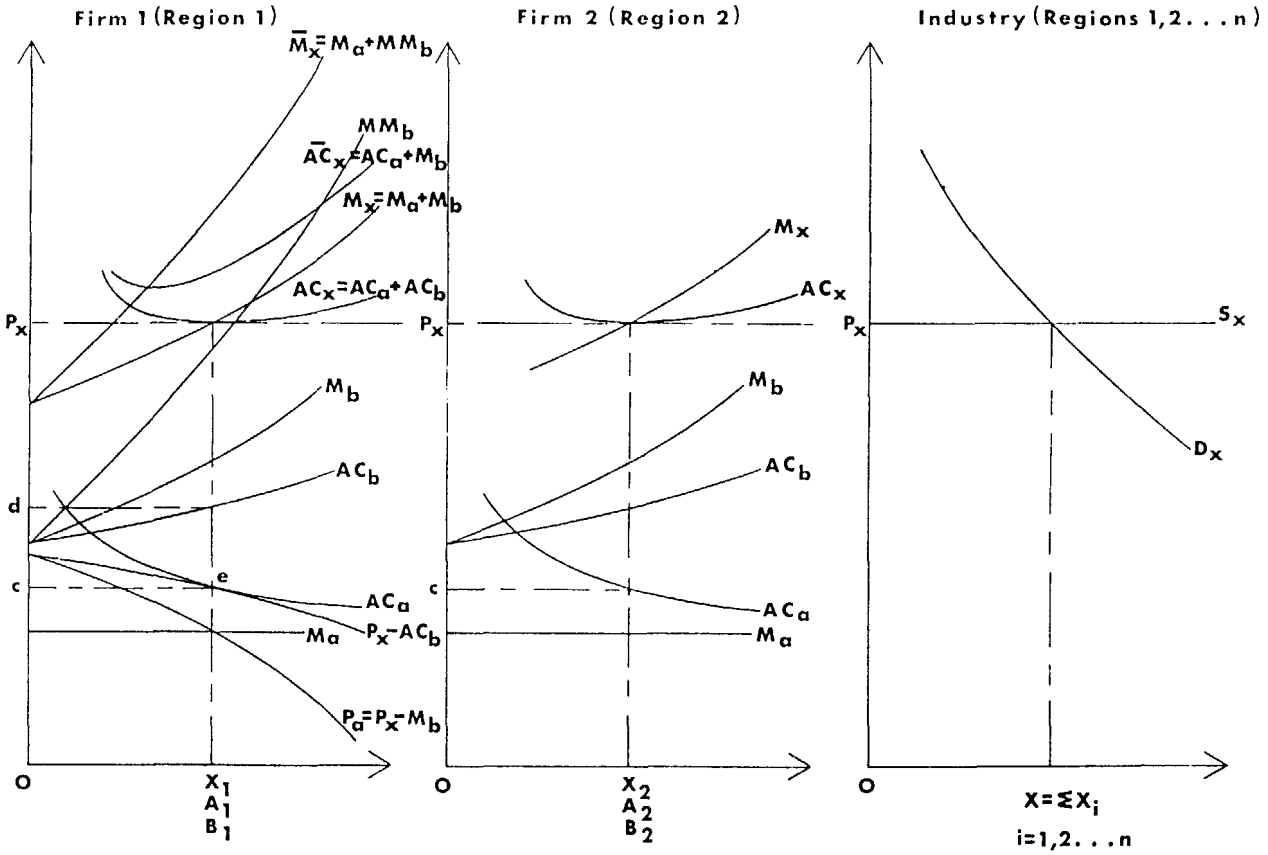

Figure 2-1. Competition With Declining Average-Cost Input

cost schedules for the two inputs and the final product, respectively, while M_a, M_b, and M_x are the corresponding marginal-cost schedules.

If the A firm in Region 1 begins to produce X, purchasing B from competitive independent firms, nothing is gained. Since the supply function for B from independent firms is given by M_b, the integrated firm will act as a monopsonist toward the B industry. Assuming that the integrated firm cannot act as a discriminating monopsonist toward the B industry, the marginal cost of B to the integrated firm is given by MM_b, the curve marginal to M_b. For the integrated firm, the average cost of producing X is $\overline{AC}_x = AC_a + M_b$, with a corresponding marginal cost of $\overline{M}_x = M_a + MM_b$. As is shown later (Chapter 4), if inputs are used in fixed proportions, the level of output will be the same after integration as before, since integration simply transforms an indirect monopsonization of B into a direct monopsonization of B. If the A and B inputs are integrated, however, output will increase. For a single X firm, producing both A and B inputs internally, the average cost of producing X will be $AC_x = AC_a + AC_b$, with a corresponding marginal cost of $M_x = M_a + M_b$.

The problem of ensuring that the firm actually sets price equal to marginal cost for the final product will still remain, however. Since there is a single input supplier, integration results in a single firm producing the final product. The monopoly problem disappears if the final product can be sold in several regions while the inputs have a regional market. This may occur, for example, if transport costs are higher for the inputs than for the final product. Thus in Figure 2.1, A, B, and X could be electricity, bauxite, and aluminum, respectively. Let us assume that neither bauxite nor electricity can be transported between regions except at some prohibitive cost, that transport costs for aluminum are negligible, that increasing returns to scale exist in electricity production, and that decreasing returns to scale exist in bauxite production. The result will be at most one aluminum producer in each region, but—if there are enough regions—the aluminum market will be competitive.

In principle the same results could be achieved under separate ownership if price-quantity agreements or discriminating monopsony is possible. The problem is that the derived-demand curve for an independent A firm in Region 1 of Figure 2-1 is given by $P_a = P_x - M_b$, which lies everywhere under AC_a. Thus no price for A exists that would allow non-negative profits for the A firm. One solution is for both A and B to be supplied under price and quantity agreements that specify A_1 units of A at price Oc, and B_1 units of B at price Od. Such an agreement effectively extracts the rent from the B industry and uses it to cover the fixed costs of producing A.

Alternatively, if a (single) X firm could act as a perfectly discriminating monopsonist toward the B industry, the derived-demand function for A would be given by $P_x - AC_b$. This new demand function is below AC_a at every point except for a tangency at point e. If the A firm now simply sets its optimal price of Oc, the downstream firm will set $AC_b + Oc = P_x$, producing X_1 units of X, and demanding A_1 and B_1 units of the two inputs.

Thus common ownership does not provide a result that *could* not be achieved through a market. But integration may be much less expensive than complex price-quantity or discriminatory arrangements. Once again, the central purpose of integration is to reduce transaction costs.

Public-Good Externalities

Public goods cause market failure because no single price exists that can be efficiently used for both the production and the distribution of a public good. While vertical integration between producer and user may ease the organizational problems created by public goods, it is horizontal integration at the user level that is critical. Thus if defense is a public good, the horizontal integration of all users in the form of a government may be required, but the government does not have to produce its own napalm. Similarly, an industry-financed research program may contract research projects out to universities. The required degree of horizontal integration, however, may cause a new market failure due to monopoly or monopsony power unless public regulation or public ownership is instituted, or the scope of horizontal integration is limited to dealing directly with the public good.[14]

One type of public good directly affected by vertical integration is information on market prices generated by unintegrated firms in a competitive intermediate-good market. Knowledge of "correct" transfer prices, provided at low cost by the market, permits better evaluation of decentralized performance within an integrated firm. In addition, the open-market price presents the option open to integrated buyers between internal and external supply, and the choice to integrated suppliers between internal and external sales. While unintegrated firms must usually provide this information at zero cost, integrated firms are under no compulsion to inform others about their internal prices.

Ideally, integrated firms should be prepared to enter the intermediate market whenever prices in that market diverge from internal marginal costs. Inter-level purchasing instructions may in practice, however, either ignore open-market prices or dictate that external

ELSEVIER Information Economics and Policy 8 (1996) 337–355

INFORMATION ECONOMICS AND POLICY

Local monopsony and free riders

David Waterman*

Department of Telecommunications, Indiana University, 515 N. Park St., Bloomington, IN 47405, USA

Received 16 January 1996; accepted 9 August 1996

Abstract

In an industry with upstream economies of scale in the distribution of differentiated products to retailers which have monopoly power within separate local market areas, the retailers have an *incentive* to exert monopsony power due to the divergence between average and marginal costs in the distribution of those inputs. The retailers increase their *ability* to exert monopsony power by forming coalitions (that is, chains) across local markets. Sufficiently large retail chains may force input price below the seller's average cost, thus 'free riding' on the level of product variety supported by other retailers. Vertical integration, cartels, or other cooperative behavior, however, can be means to control the level of product variety, and may increase both industry profits and economic welfare. Policy applications to the cable television, motion picture, and pharmaceutical industries are discussed.

Key words: Monopsony; Vertical integration; Cable television; Motion pictures; Pharmaceutical

JEL Classification: D42; L12; L82

1. Introduction

Accompanying recent growth of the cable television industry has been marked change in the industry's ownership structure. From 1985 to 1995, the national market shares of the four largest multiple cable television system operators (MSOs) increased from 24.9% to 54.6% (61.3% including announced transactions)

* Corresponding author: Tel: 812-855-6170; fax: 812-855-3125; e-mail: waterman@indiana.edu.

PII S0167-6245(96)00021-2

of U.S. basic cable subscribers (FCC, 1990, Appendix G; FCC, 1995, Appendix G.) Vertical ties between MSOs and cable programming networks have also become widespread. Of 129 nationally distributed networks that the FCC identified in November, 1995, 66 had ownership ties to cable system operators, including 11 of the 15 most widely distributed basic cable networks, and 4 of the 6 largest premium networks. An interesting feature of vertical relationships in cable is that many involve 'equity sharing' arrangements in which two or more MSOs each have partial ownership of a single cable network; MSOs shared equity in 23 of the 66 integrated networks in 1995 (FCC, 1995, Appendix H).

These horizontal and vertical ownership ties have attracted policy scrutiny, especially those of the two largest MSOs, Telecommunications, Inc. (TCI) and Time-Warner, which respectively accounted for 26% and 16% of U.S. cable subscribers in November, 1995 (29% and 19% including announced transactions), and had 5% or greater ownership interests in 38 and 18 nationally distributed cable networks (FCC, 1995, Appendix G, H). As mandated by the 1992 *Cable Television Consumer Protection and Competition Act*,[1] the FCC set a limit of 30% on the proportion of U.S. homes passed by cable that can be accounted for by a single MSO and a limit of 40% on the proportion of a cable system's channels which the system can fill with programming in which it has an equity interest (FCC, 1993).[2] The Justice Department and the Federal Trade Commission have also investigated recent horizontal and vertical merger activity in the cable industry.[3]

One aim of this paper is to better explain the motives behind these and similar structural developments in other industries, and to assess their welfare consequences. I hypothesize that an incentive for formation of horizontal coalitions such as MSOs or movie theater chains may be to exert monopsony power with product suppliers upstream (that is, cable networks or movie producer/distributors). The theoretical model I develop also suggests that vertical integration, cartels, or other forms of industry-wide cooperation may be means to limit detrimental effects which the exercise of monopsony power may have on aggregate industry profits, and that such 'solutions' to monopsony power may be welfare increasing.

A second aim of the paper is to contribute to the economic theory of monopsony, and in that process, to identify misdirected public policies that have resulted from incorrect application of monopsony theory. In the standard textbook treatment, monopsony is a 'flip side'[4] version of monopoly. A firm's incentive to exercise monopsony power depends on an input supply curve which slopes upward

[1] Cable television consumer and competition protection act, Pub. L. No. 102-385, 106 Stat. 1460, 1992, codified at 47 U.S.C. 521–55.

[2] The FCC's 30% limit was struck down by a lower court, and at this writing is pending appellate review (FCC, 1994, par. 140).

[3] Andrews (1995); FCC (1995, Appendix G); FCC (1994, Appendix G); Robichaux (1995).

[4] Carlton and Perloff (1990) use this term.

because additional inputs can be attracted into the market only at increasing marginal prices. Analogous to the monopolist's incentive to restrict output because of downward sloping demand, the monopsonist restricts input purchases because it considers the higher prices it must pay for all inframarginal input units. In the model of this paper, a firm's incentive to exercise monopsony power has a different origin, and has policy implications which depart from those of the 'flip side' model.

In the present model, competing upstream suppliers manufacture differentiated products (the inputs) under constant returns to scale, but they distribute these inputs under conditions of increasing returns with respect to the number of downstream firms that buy them. The downstream firms are geographically separated local retailers with monopoly power. These retailers simply offer the differentiated products they buy from upstream to consumers; they are, that is, simultaneously both monopolists and monopsonists at the local level. The mass media are prominent among industries that appear to have these characteristics. Cable programming networks, for example, incur a 'first copy' cost in the creation of their programming, but those programs can be electronically distributed by satellite to additional local cable systems by little more than the flip of a switch; thus the economies in input distribution. Cable systems typically enjoy monopolies of cable service within local market areas; as input buyers, they provide the only practical outlet for large numbers of television productions.[5]

The basic incentive to exercise monopsony power in this model arises because upstream economies of scale in distribution lead to a divergence between marginal and average costs at the input level. A localized downstream buyer would like to exercise monopsony power to force it's input price near to the supplier's marginal cost of distribution while still enjoying the product variety created by an upstream industry selling at average cost to downstream buyers in all other local markets. The successful exercise of monopsony power by this downstream buyer necessarily reduces the equilibrium amount of product variety supplied from upstream, because suppliers exit the industry in response. This reduction in variety occurs only in proportion to the downstream monopsonist's share of the national market, however, permitting it to substantially 'free ride' on contributions to upstream suppliers' fixed costs made by other downstream firms.

An important distinction made in this paper is that between the *incentive* and the *ability* of a firm to exercise monopsony power at the local level. While the downstream coalition's incentive to exert monopsony power follows as described from the upstream cost conditions it faces, that power only materializes in this

[5] Although several competitors to cable systems exist, including Direct Broadcast Satellite (DBS) operators, Multichannel Multipoint Distribution Systems (MMDS), Satellite Master Antenna Systems (SMATV), and Home Satellite Dishes (HSD), and 'overbuild' cable systems have entered several local markets, the national market shares of the 'multichannel video programming distribution' market aggregated to under 9% as of September, 1995 (FCC, 1995, Appendix G, Table 1).

model with the bargaining leverage that comes from having a substantial national market share.

A failure to distinguish clearly between the incentive and the ability to exercise monopsony power has led to controversy in the literature and confusion among policymakers. Matthewson and Winter (1987) refer to the presumption that a retailer can exert monopsony power with its supplier simply because it has the only outlet for a given product within a local area as "intuitive, popular, and wrong" (p. 1058), but they offer no further explanation. The general notion that a local monopoly retailer with a relatively small fraction of the national market would have relatively little buying power appears to underlie the early 'countervailing power' hypothesis of Galbraith (1952), as well as Matthewson and Winter's assertion about monopsony. As the latter authors note, however, no analytic basis for the hypothesis that both local *and* national market shares are relevant to monopsony power has been offered. I provide this basis with a simple bilateral bargaining model. I further argue that failure to recognize the importance of both local and national market shares of buyers has led regulatory and antitrust authorities to misjudge the threat of monopsony power in cable television and perhaps other industries.

There is an extensive general literature on monopsony and its relationship to vertical integration, but this literature mostly focusses on the theoretical effects of monopsony on efficient vertical contracting. For surveys, see Scherer and Ross (1990, ch. 14), Perry (1989) and Blair and Harrison (1993). Blair and Harrison offer a brief graphical treatment of monopsony and economies of scale and they analyze a taxonomy of antitrust cases involving monopsony. Previous authors, however, do not treat geographically localized monopsony, or the effects of monopsony on product variety, in an explicit or substantial way. The significance of monopsony power has been recognized in previous studies of some mass media industries, notably by Conant (1960) in the case of the theatrical motion picture industry and by Besen and Johnson (1984) in the case of broadcast television. These authors, however, do not explicitly consider how chain ownership across different local markets might affect monopsony power or product variety.

I begin in Sections 2.1 and 2.2 below by setting out the parameters of a basic one period bargaining model between upstream and downstream firms. This model defines the incentives of both upstream and downstream firms to accumulate bargaining power in the input market by forming horizontal coalitions. Then in Sections 2.3 and 2.4, I develop alternative comparative statics models which define the endpoints of the input price contract curve, or bargaining set.

The focus of the analysis is on downstream coalitions. One end of the contract curve is defined by the input 'price taker' retailer model, in which the reservation input price of downstream coalitions is established as the level at which all of their local monopoly profits are extracted by upstream suppliers. The other end of the contract curve is defined by a downstream 'single price maker' model that establishes the reservation input prices of upstream suppliers. In that model, a

single retailer coalition forces input prices near to marginal costs of distribution, reducing the supply of differentiated products produced upstream. The incentives for this downstream coalition to exert monopsony power diminish as it becomes larger, however, because the coalition internalizes a larger proportion of the negative externality which its price making behavior has on the supply of differentiated products available to all markets. The balance of these opposing forces as a retailer coalition becomes larger – increasing power to exert monopsony power, but decreasing incentives to do so – determines the extent to which such power will actually be exerted by downstream firms, and thus the equilibrium level of product variety. In Section 2.5, an example illustrates the effects of the various bargaining outcomes on product variety in the 'price taker' and 'single price maker' models.

In Section 2.6, the destructive effects on product variety arising from myopic behavior by downstream retailer coalitions throughout the industry are first shown. How vertical integration or cooperative behavior could increase industry profits by moving product variety back toward the industry profit maximizing level is then discussed (Section 2.7). As a next step (Section 2.8), I make economic welfare comparisons to show whether consumers are better or worse off due to the formation of horizontal coalitions, vertical coalitions, or cooperative behavior. Finally in Section 3, I summarize and discuss implications for antitrust and regulatory enforcement in the media and also in the pharmaceuticals industry.

2. The model

2.1. Basic assumptions

There are M local markets of equal size with symmetric demand conditions, $i = 1, \ldots, M$. In each one of these markets, a single monopoly retailer markets a menu of up to N differentiated products supplied to it by upstream firms, $j = 1, \ldots, N$. For the downstream retailer in market i:

$$\Pi_i = \sum_{j=1}^{N} [q_{ij}(p_{ij} - c) - Z_{ij}] \tag{1}$$

where q is the number of buyers per product offered, p is final price for each sale, and c is a constant marginal cost of retail distribution. The term Z represents a negotiated lump sum to be paid back to suppliers from retail revenues. Thus there are no transactions costs and no marginal component to the input price contract.

There are N upstream producer/distributors, one for each product. For each of these firms:

$$\Pi_j = \sum_{i=1}^{M} Z_{ij} - K \tag{2}$$

where K is a fixed cost of producing each product, which is assumed constant across products. Marginal costs of upstream distribution are zero.

As a result of entry and exit in the upstream industry, N is determined as an equilibrium condition of the model. For simplicity, I assume that each of these products is equally attractive to consumers. That is, the N products are always symmetrically distributed in some product space. No entry is permitted in the downstream market.

Demand is defined directly, $q_{ij} = q_{ij}(p_{ij}, p_{i,-j}, N)$ where the subscript, $-j$, indicates the vector of prices of all products except j. $p_{ij} = \infty$ for all services not in the market. $\partial q_{ij}/\partial p_{ij} < 0$; $\partial q_{ij}/\partial p_{i,-j} > 0$; $\partial q_{ij}/\partial N < 0$; and $\partial^2 q_{ij}/\partial N^2 > 0$. The latter two derivatives indicate that all products are substitutes, but that demand for an individual product decreases with an increase in variety, but at a decreasing rate.

2.2. *Bargaining power and horizontal coalitions*

I now describe a simplified one period input price bargaining process between upstream suppliers and downstream retailers. At the beginning of the period, there is simultaneous negotiation across the nation between suppliers and retailers for all potentially available products in all local markets. There is no uncertainty about final demand and there is complete information about the reservation prices of all parties. As noted above, Eq. (1) and Eq. (2) reflect zero bargaining costs. Based on results of the bargaining, upstream firms decide whether or not to produce, and downstream firms decide which products, if any, to offer to consumers. During the period, retail transactions take place and settlements between producers and retailers are made. The entire process is then repeated in the next period.

Note that since there is no marginal input price component, double marginalization is not involved in this model. That is, settlements are made in terms of lump sum Z's without a priori uncertainty of what final demand will be. However, alternative equilibria can be more usefully compared if we consider the negotiations to actually take place in terms of another variable, r, $0 < r < 1$, the percentage share of total retail revenues which will accrue to either party after transactions are completed. That is,

$$r_{ij} = \frac{Z_{ij}}{p_{ij} q_{ij}}. \tag{3}$$

Horizontal coalitions may be formed among upstream or downstream firms for the purpose of increasing those firms' bargaining power in the input market. Local monopolists may combine across local markets into 'chain coalitions', which may be of different sizes (the largest possible would combine all local markets into one national firm). Let m_d, $d = 1, \ldots, D$, define the sizes of downstream chain coalitions, so that m_d/M measures the proportion of the national market controlled

by the dth coalition. For tractability, upstream coalitions are restricted to be of equal, that is, symmetric, sizes. Let n be the size of the representative supplier coalition, so that n/N is the proportion of all differentiated products controlled by each of the upstream coalitions.

If a supplier coalition fails to make a sale to the dth coalition, that supplier coalition can receive no revenues from m_d/M of the national market. Comparably, a downstream coalition risks the increment to its retail revenues which the products controlled by the nth coalition contribute.[6] This circumstance suggests a range over which the input price contracts, i.e., the r_{ij}s, may lie. As Eq. (2), Eq. (3), and the assumption of upstream entry and exit imply, the r_{ij}s in turn determine equilibrium N.

What determines the point along the relevant contract curve at which a deal negotiated between a given downstream and upstream coalition will be transacted? I hypothesize the following general solution:

$$r^e = r^{\min} + g(r^{\max} - r^{\min}) \tag{4}$$

where the subscript e, indicates the equilibrium solution, the superscript 'min' indicates the reservation price of the downstream coalition in the bargain, and 'max' indicates reservation price for the upstream firm. Define $g = g(m_d/M, n/N)$, such that $0 < g < 1$ and $\partial g/\partial(m_d/M) < 0$ and $\partial g/\partial(n/N) > 0$. That is, bargaining power is determined by the relative national market shares of the upstream and downstream coalitions.

The function (4) reflects a central postulate of cooperative game theory: that relative bargaining power in a bilateral game is inversely related to how much either party has to lose if no deal is struck. As m_d rises from 0 to 1, the proportion of the upstream supplier's total revenues at risk in the bargain increases linearly, while those of the retailer coalition remain constant. Conversely, as n/N increases, the retailer coalition's proportion of revenues at risk increases at an increasing rate, depending on $\partial q_{ij}/\partial N$ *and* $\partial^2 q/\partial N^2$, while those of the upstream coalition remain constant. An implication of Eq. (4) is thus that even though a downstream retailer is by definition a monopsonist as well as a monopolist within its local market area, that retailer may exercise negligible monopsony power with sellers if it accounts for a negligible proportion of the sellers' national market.

The next step is to define $r^{\max}$ and $r^{\min}$, which are equivalent to the reservation input prices of the downstream and upstream coalitions, respectively, in any given bargaining game. As will be seen, $r^{\min}$ varies directly with m_d because the larger is

[6] Even when upstream entry is possible, as I assume to be the case, successful entrants must attract revenues from all local markets combined to cover fixed production costs plus prevailing profit margins. An individual local monopsonist thus cannot anticipate that if a bargain with one potential supplier potential fails, the incremental reduction in the number of products it offers to consumers will be made up by a new supplier during that period.

m_d, the greater is the negative effect that input price setting by the retailer coalition would have on N, and thus on the product variety which coalition m_d can offer its local consumers. I consider two extreme alternative models in order to identify the end points of these contract curves in terms of r. In the 'price taker' model immediately below, equilibrium r is at the reservation level of the downstream coalition, r^{max}. In the 'single price maker' model, r is at the reservation level of the upstream coalition, r^{min}.

2.3. Price taking retailer coalitions

For each downstream coalition,

$$\Pi_{m_d} = \sum_{i=1}^{m_d} \Pi_i. \tag{5}$$

For the representative upstream coalition,

$$\Pi_n = \sum_{j=1}^{n} \Pi_j. \tag{6}$$

Downstream firms maximize profit w.r.t. all p_{ij}s within m_d. To find equilibrium solutions, I apply symmetry across and within markets and set Eq. (5)=0. That is, the downstream firms behave as monopolists in the final market, but all their revenues above costs are extracted by the upstream firms. Further let Eq. (6)=δ>0 for upstream coalitions. That is, some level of excess profit per product may be earned by upstream coalitions before entry occurs. Maximization yields the following first order conditions:

$$rpQM = (K + \delta)N, \tag{7}$$

$$(p - c)\frac{\partial Q}{\partial P} + Q = 0, \tag{8}$$

$$(1 - r)p - c = 0, \tag{9}$$

which are three equations in p, N and r.

In order to solve this and subsequent systems, I specify the following specific aggregate demand function for each local market area:

$$Q_i = \sum_{j=1}^{N} q_{ij} = (J - \alpha p_i)N^{\beta} \tag{10}$$

where $J, \alpha > 0$, and $0 < \beta < 1$. The parameter β thus measures the elasticity of aggregate demand with respect to product variety. Consistent with the general demand function above, $\partial Q_i/\partial N > 0$, and $\partial^2 Q_i/\partial N^2 < 0$.

Using (10), the solutions are:

$$p^{pt} = \frac{J + \alpha c}{2\alpha}, \tag{11}$$

$$r^{pt} = \frac{p^{pt} - c}{p^{pt}}, \tag{12}$$

$$N^{pt} = \left[\frac{(p^{pt} - c)(J - \alpha p^{pt})}{K + \delta}\right]^{\frac{1}{1-\beta}}, \tag{13}$$

where 'pt' indicates the price taker solution.

Note from (11) that final prices are independent of product variety. Reflecting the condition of zero profits downstream, the equilibrium division of revenues, r^{pt}, is equal to the per-final-sale markup over retailer marginal costs. The equilibrium number of products, N^{pt}, is increasing in β, the elasticity of consumer demand w.r.t. product variety, and decreasing in K, production costs.

2.4. Price making retailer coalitions

To establish the other end of the contract curve, imagine that one retail chain coalition, m_1, manages to challenge the price making behavior of upstream firms, while other downstream coalitions remain as price takers. All downstream coalitions maximize w.r.t retail prices as before. As an input price maker, however, m_1 also controls theZ_{ij}s relevant to the markets in its coalition.

Two further assumptions are introduced at this point. The first is that successful bargains are actually struck for all available products in all local markets at some set of input prices. That is, $N_{m_1} = N_{-m_1} = N$, where $-m_1$ indicates all other price taking retailer coalitions. The second assumption is that the price making coalition makes symmetric bargains with supplier coalitions in all markets which it controls; that is, a given downstream coalition makes bargains with each supplier coalition on the assumption that the same bargain will be made with all other supplier coalitions.

The following Lagrangian objective function can then be written for m_1:

$$L_{m_1} = m_1[(p_{m_1} - c)Q_{m_1} - Z_{m_1}] + \lambda N[(m_1 Z_{m_1} + (M - m_1)Z_{m_1}) - (K + \delta)]. \tag{14}$$

That is, the downstream coalition maximizes profits subject to the constraint that all upstream firms or coalitions must at least earn δ per product.

Maximization of Eq. (14) w.r.t. p_{m1}, Z_{m1}, λ, and N yields $\lambda = 1$ and three other equations in five unknowns, p_{m_1}, p_{-m_1}, r_{m_1}, r_{-m_1}, and N. There is no reason for price making behavior in one market to affect the equilibrium in other price taking local markets, so Eq. (8) and Eq. (9) can be applied to the $M - m_1$ price taking markets. Using Eq. (10), these five equations solve to the following:

$$p^{spm}_{m_1} = p^{spm}_{m_1} = p^{pi} = p^* = \frac{J - \alpha c}{2\alpha}. \tag{15}$$

$$r^{spm}_{m_1} = \beta\frac{m_1}{M}\left[\frac{p^* - c}{p^*}\right]. \tag{16}$$

$$r^{spm}_{m_1} = r^{pi} = \frac{p^* - c}{p^*}. \tag{17}$$

$$N^{spm} = \left[\frac{\left(\beta\frac{m_1}{M} + \frac{M - m_1}{M}\right)(p^* - c)(J - \alpha p^*)}{K + \delta}\right]^{1/(1-\beta)} \tag{18}$$

where 'spm' indicates the single price maker solution.

Comparing these results with those of the price taker model, note that retail prices are independent of N and are unchanged. Equilibrium retail prices are indicated by p^* in this and all models henceforth, because they do not vary in any of the cases we consider. However, $r^{spm}_{m_1}$ falls to a fraction of r^{pi} and $N^{spm} < N^{pi}$, these differences depending on the magnitudes of β and m_1/M. The price making coalition's actions have a negative externality effect on product variety. That is, the retailer coalition behaves myopically, considering only the relatively marginal impact of its setting of r on the supply of differentiated products which can be made available by upstream firms.

If m_1 is very small, then $r^{spm}_{m_1}$ goes to zero. That is, the actions of a price making retailer coalition having a very small national market share will have a negligible effect on product variety. At the other extreme, if $m_1 = M$, then $r^{spm} = \beta r^{pi}$ and $N^{pm} = \beta^{\beta/(1-\beta)} N^{pi}$. The latter solutions for r and N represents those of a price making nationwide retailer chain coalition; the externality problem of local myopic behavior thus disappears because the retailer suffers the full effects of its input pricing behavior.

The downstream coalition with a small fraction of the national market thus perceives a relatively inelastic supply of differentiated products w.r.t. r. This supply function becomes flatter, however, as the downstream coalition's national market share increases. Reflecting the assumption of constant costs, K, in upstream input production, the supply curve faced by the single national retailer coalition becomes perfectly flat. For this reason, the national retailer no longer has an incentive to exert monopsony power. Equilibrium product variety bought from upstream suppliers and offered to consumers by the national price making downstream retailer coalition is below that of the price taker model, not due to monopsony power, but because the national firm can now coordinate a monopolistic restriction of product variety at the output level. A monopolist's incentive to restrict product variety has been shown by White (1977), Mussa and Rosen (1978), and other later authors.

In terms of r, the extreme points on a set of contract curves which retailer and

supplier coalitions face in their negotiations over input price are represented by (12) and (16). Note that while this model is basically set out as a cooperative game, there are differences with the usual formulation. First, while the upper limit of the contract curve, $r = (p^* - c)/p^*$, is also the retailer coalition's threat point, the lower limit of the contract curve, $\beta m_1/M[(p - c)/p^*]$, is not the threat point of an individual supplier coalition in the usual sense. That is, if no deal is struck, the upstream coalition effectively realizes a division of profits equivalent to $r = 0$ (its threat point). However, it would be irrational for the retailer coalition to force r below $\beta m_1/M[(p^* - c)/p^*]$, given the assumption that identical deals are struck with all other supplier coalitions.

A second difference is that while any bargain in r must be locally Pareto superior for the two parties to find it in their interest, any bargain in r below $(p^* - c)/p^*$ is not necessarily Pareto superior at the national market level (unless $m_1 = M$). The latter result may occur due to the negative externality on product variety which affects retailers in other markets.

2.5. The bargaining outcome illustrated

The extreme points and other potential possible bargaining outcomes along the contract curve are illustrated in Fig. 1 for the case in which $(p^* - c)/p^* = 0.8$ and $\beta = 0.5$. Reflecting the solution of the price maker model, the reservation price of the representative upstream coalition, and thus the lower limit of r, increases with m_1/M, as indicated by the line AB. The line CD shows the reservation price of any given sized retailer coalition; this reservation price does not vary with m_1/M due to the assumption of constant returns to scale downstream. The relevant contract curve is thus a vertical line between CD and AB, intersecting AB at the appropriate value of m_1/M.

It can easily be shown that point B in Fig. 1, where the downstream coalition has 100% of the national market and extracts all revenues over costs from upstream firms, is the industry joint profit maximization point for r, and thus N. Of particular interest, it is evident from Fig. 1 that the actual bargaining outcome, r^c, could fall above or below this point. A specific logit function for g demonstrates. Let

$$g = \frac{2}{1 + e^{-x}} - 1, \quad \text{where} \quad x = \frac{n/N}{m_1/M}. \tag{19}$$

As x goes to 0, g goes to zero. As x becomes large, g goes to 1. The function g thus varies monotonically between 0 and 1, depending on the relative market share of the upstream and downstream firms as hypothesized in Eq. (4). The family of curved lines in Fig. 1 illustrates the resulting bargaining outcome for several alternative given values of n/N, the representative supplier coalition's national market share. In cases where market shares are relatively low upstream (for example, where $n/N < 0.2$), r^c falls below B for some values of m_1/M.

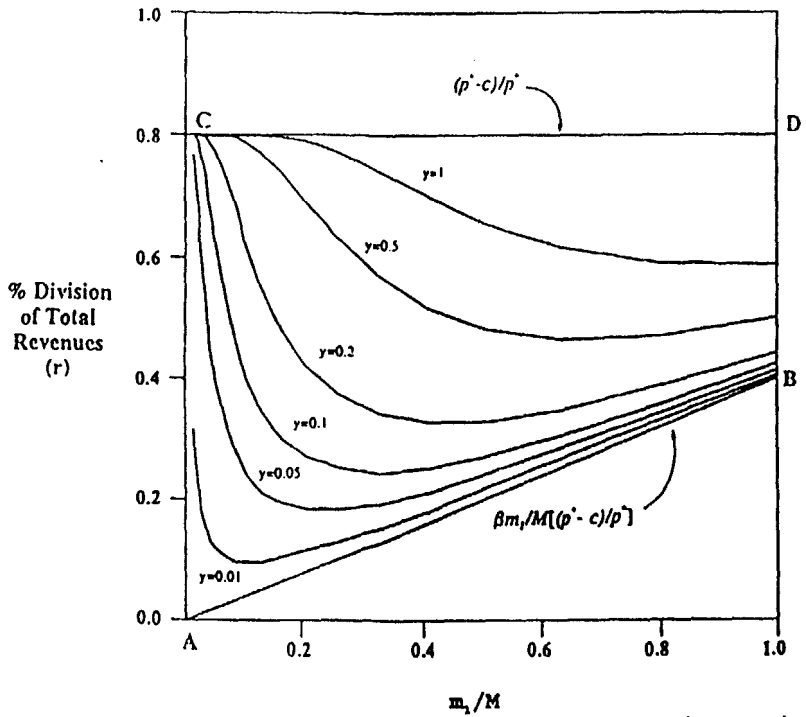

Fig. 1. Equilibrium input prices: % division of revenues vs. national market share of representative retailer coalition for various national market shares of representative supplier coalitions ($y = n/N$).

The Fig. 1 illustration is, of course, just that. The model nevertheless demonstrates that myopic behavior of downstream firms may leave product variety below the industry profit maximizing level.

2.6. Industry-wide price making behavior

If price making behavior were successfully practised by all of many local monopoly retailer coalitions, rather than just one, the cumulative effects of this myopic behavior could essentially shut down the industry. If all retailer coalitions are of size m_1 and practice price making behavior, then

$$r^{apm} = \left[\frac{m_1}{M-(M-m_1)\beta}\right]^{1/(1-\beta)} \beta\left(\frac{p^*-c}{p^*}\right), \tag{20}$$

$$N^{apm} = \left[\frac{(m_1\beta/[M-(M-m_1)\beta])(p^*-c)(J-\alpha p^*)}{K+\delta}\right]^{1/(1-\beta)} \tag{21}$$

where 'apm' indicates the 'all price maker' retailer solution. Both of these expressions go to zero as m_1 goes to 0.

2.7. Vertical integration and cooperative behavior

The ability of individual downstream retailer coalitions to myopically reduce input prices below the industry's joint profit maximizing level creates an incentive for firms to form structural or other relationships which limit the externality effects of that behavior. It was noted above that one way to resolve the myopia problem is the formation of downstream horizontal coalitions so large that they substantially internalize the coalition's incentive to impose a negative externality on other retailers.

It is also easily shown that a vertically integrated national coalition of all retailers and suppliers produces the industry profit maximizing level of product variety (at B in Fig. 1) by implicitly setting $r^c = \beta(p^* - c)/p^*$. This result would be achieved, of course, if the entire industry had a common owner, or if there were an industry-wide cartel among separately owned firms.

2.8. Welfare analysis

Since optimal retail price does not vary with product variety in the models presented, product variety is the only variable affecting economic welfare. We can therefore compare the welfare outcomes of the various comparative statics models simply in terms of optimal N. Total welfare, the sum of producer's surplus (net industry profits) and consumer's surplus is defined as

$$W = M(p-c)Q - NK + \int_{p^*}^{J/2} (J-\alpha p)N^{\beta}\,dp. \tag{22}$$

Maximizing w.r.t. N yields

$$N^{w} = \left[\frac{\beta(p^*-c)(J-\alpha p^*) + (J-\alpha p^*)^2/2\alpha}{K}\right]^{1/(1-\beta)}. \tag{23}$$

As would be expected, welfare increases in β and falls in K.

Comparison with the price taker, single price maker, industry profit maximizing, and all price maker industry equilibria for product variety derived above, yields:

$$N^{apm} < N^{c} < N^{w}, \tag{24}$$

$$N^{apm} < N^{c} < N^{spm} < N^{pt}, \tag{25}$$

$$N^{w} \lesseqgtr N^{spm}; \quad N^{w} \lesseqgtr N^{pt},$$

where N^c indicates the cartel, or industry profit maximizing equilibrium obtained

by setting $m_1 = M$ in Eq. (18). Whether N^w is greater or less than N^{spm} or N^{pl} depends on the parameters of the demand function: J, α, and β; on the amount of excess profits earned upstream, δ; and for the case of N^{spm}, on the national market share of the single price maker retail coalition.

Based on the assumed demand function (4), these results thus show that industry-wide coordination, as represented by N^c, unambiguously improves welfare over the destructive 'all price maker' case, represented by N^{apm}. Both the N^c and N^{apm} cases, however, leave product variety unambiguously below the welfare optimum. Cartel behavior in this model is thus helpful, but not ideal, from the public viewpoint.

The ambiguous relationship of N^w to N^{spm} and N^{pl}, is to be expected since the total amount of producers' and consumers' surplus necessarily depends on specific parameters of the demand function. In fact, it is well-known that optimal product variety depends more generally on the form of the demand function (Spence, 1976; Dixit and Stiglitz, 1977; Tirole, 1988). The welfare results in Eq. (24) are thus not necessarily robust to alternative specifications of Eq. (4).

The welfare results using Eq. (4) are nevertheless a reasonable example of how vertical integration or collusion can improve economic welfare even in the absence of transactions cost savings. Note also that even though the successful exercise of myopic monopsony power on product diversity can reach anticompetitive levels, its exercise could improve welfare within a certain range, as could the countervailing exercise of bargaining power upstream. In the case of media industries such as cable television, one might also argue on non-economic grounds that there is a social value to high product diversity.

3. Summary and policy discussion

In an industry with upstream economies of scale in the distribution of differentiated products to retailers which have monpopoly power within separate geographic areas, the retailers have an incentive to exert monopsony power. Unlike the standard textbook model, the firm's incentive is to exploit the difference between average and marginal costs in the distribution of those inputs, in order to free ride on the level of product variety supported by other downstream firms. This incentive to exercise monopsony power does *not* imply, however, the ability to do. To gain that ability, downstream retailers form coalitions across local markets.

Successful exertion of monopsony power by downstream coalitions having less than 100% of the national market may reduce product variety below industry profit maximizing levels. Economic welfare may rise or fall, depending on the optimum product variety, but if monopsony power is exerted beyond a certain level, welfare unambiguously falls. The model suggests that vertical integration or industry-wide cooperative behavior can serve to internalize the negative externality of myopic

input pricing behavior, returning the industry toward a joint profit maximizing equilibrium. Such coalitions may increase consumer welfare by returning product variety toward the welfare optimum.

The free rider model suggests one rationale for recent structural developments in the cable television industry. Horizontal growth by the larger MSOs may be attempts to exercise monopsony power with programming suppliers. Chipty (1995) reports econometric evidence suggesting that larger MSOs receive substantial discounts from programming suppliers due to the exertion of monopsony power. Large and widely acknowledged differentials between the (relatively low) licensing fees that larger MSOs have paid for basic and premium cable programming networks and the (relatively high) fees paid by smaller 'independent' cable operators, and 'wireless' cable operators for the same networks are also consistent with the monopsony hypothesis (National Telecommunications and Information Administration, 1988; Waterman and Weiss, 1996).

The extensive vertical integration into programming by the largest MSOs, and more generally, the common practice of 'equity sharing' in networks by MSOs, may be attempts to internalize the negative externality which opportunistic input price setting by larger MSOs creates in the absence of integration. Of course, control of 25–30% of the national market by the leading firm (TCI) may not seem excessive. In the presence of upstream economies of scale in cable networking, however, the bargaining model suggests that substantial monopsony power over programming suppliers could be exerted by such a firm.

In defending its choice of 30% of U.S. cable homes passed for the size limit on MSOs, the FCC has argued that although local cable systems typically have nearly 100% local market shares, the national market concentration of MSOs, based on their shares of all U.S. cable subscribers and as measured by the HHI, was at or below the Justice Department's minimum '1000' standard ordinarily warranting investigation in horizontal merger cases (FCC, 1994, 1995; see also FCC, 1990). The FCC is simply wrong to interpret an HHI measure in this way. As the *1992 Horizontal Merger Guidelines* (United States Department of Justice and Federal Trade Commission, 1992) make clear, the HHI standards are concerned with the accretion of market power through unilateral or coordinated behavior that would result from a merger within a particular market within which other firms compete for the *same* customers (or inputs). Obviously, however, there is no national market for cable subscriberships. A similar critique applies to rules of thumb about the relationship between market power and the national market share of a single firm (e.g., that a firm having less than 35 or 40% of the market is unlikely to have excessive market power). Such rules of thumb were relied upon by many commenters in the FCC proceedings to argue for an MSO size limit of 40% or more. The appropriate criteria for assessing monopsony power in such cases is the relative bargaining power of the MSO and the various program suppliers, which in turn depends on the extent of upstream economies of scale and alternative means of distribution which the program suppliers may have.

case in *Syufy* was weak. As the Appeals Court decision at one point referred to Syufy, he was 'a relatively tiny regional entrepreneur' (pp. 63, 576), having in 1986 a 1.3% national market share of theater screens and apparently no significant holdings in major markets other than Las Vegas (compared to a 9.1% national market share for United Artists, the largest theater chain in the U.S.) (Motion Picture Association of America, 1990; Variety, January 14, 1987). The suggestion of the free rider model is that control of the Las Vegas market was in itself simply not a formidable threat to hang over the heads of theatrical distributors, who received about 99% of their revenues from other sources. As one of the distributors, James Spitz, testified at the trial, "...if he [Syufy] would have pressed, if he would have come to Jimmie Spitz and said, 'I'm not going to pay you this percentage for the film'", I would have said, "Fine, Ray, we'll just stay out of the [Las Vegas] marketplace". (pp. 63, 582).

In short, both local *and* national market shares are relevant to the exercise of monopsony power. Antitrust and regulatory policies must recognize this distinction explicitly.

Acknowledgments

I am grateful to Timothy J. Brennan, Thomas P. Lyon, Sheldon Kimmel, Eric Rasmusen, and Lawrence J. White for comments. Errors are my own. Earlier versions of this paper were presented at the Fifth Annual Conference of the American Law and Economics Association, Berkeley, CA, May 12–13, 1995, and at the Second Annual Conference of the Consortium for Research on Telecommunications Policy, Evanston, IL, May 10–11, 1996.

References

Andrews, E.L., 1995, After delay, FTC allows QVC's takeover, New York Times, February 4, C17.
Besen, S. and L. Johnson, 1984, An analysis of the Federal Communications Commission's group ownership rules, Rand report no. N-2097-MF (Rand, Santa Monica, CA).
Blair, R.D. and J.L. Harrison, 1993, Monopsony (Princeton University Press, Princeton, NJ).
Carlton, D.W. and J.M. Perloff, 1990, Modern industrial organization (Little Brown, Glenview, IL).
Caves, R.E., M.D. Whinston and M.A. Hurwicz, 1991, Patent expiration, entry, and competition in the U.S. pharmaceutical industry, Brookings Papers on Economic Activity, annual.
Chipty, T., 1995, Horizontal integration for bargaining power: Evidence from the cable television industry, Journal of Economics and Management Strategy 4, no. 2, 375–397.
Conant, M., 1960, Antitrust in the motion picture industry: An economic and legal analysis (University of California Press, Berkeley, CA).
Dixit, A.K. and J.E. Stiglitz, 1977, Monopolistic competition and optimal product diversity, American Economic Review 67, 297–308.

FCC (Federal Communications Commission), 1990, Competition rate deregulation and the commission's policies relating to the provision of cable television service, F.C.C. docket no. 89-600, Report no. 90-276 (Federal Communications Commission, Washington, DC).
FCC (Federal Communications Commission), 1993, Implementation of Sections 11 and 13 of the cable television consumer protection and competition act of 1992, Horizontal and vertical ownership limits, cross-ownership limitations and anti-trafficking provisions, Second report and order, MM docket no. 92-264, Released October 23, 1993 (Federal Communications Commission, Washington, DC).
FCC (Federal Communications Commission), 1994, Annual assessment of the status of competition in the market for the delivery of video programming, Implementation of Section 19 of the cable television consumer protection and competition act of 1992, CS docket no. 94-4, Released September 28, 1994 (Federal Communications Commission, Washington, DC).
FCC (Federal Communications Commission), 1995, Annual assessment of the status of competition in the market for the delivery of video programming, Implementation of section 19 of the cable television consumer protection and competition act of 1992, CS Docket No. 95-61, Released December 11, 1995 (Federal Communications Commission, Washington, DC).
France, M., 1996, Cake for lawyers, crumbs for clients, Business Week, April 29, 86.
Frech, H., III, 1978, Market power in health insurance, effects on insurance, and medical markets, Journal of Industrial Economics 27, 55–72.
Galbraith, J.K., 1952, American capitalism: The concept of countervailing power (Houghton Mifflin, Boston, MA).
Hampton, B., 1970, History of the American film industry: From its beginnings to 1933 (Dover, New York).
Lewis, H., 1933, The motion picture industry (Van Nostrand, New York).
Matthewson, G.F. and R.A. Winter, 1987, The competitive effects of vertical agreements: Comment, American Economic Review 77, no. 5, 1057–1062.
Motion Picture Association of America, 1990, U.S. Economic Review (Motion Picture Association of America, Encino, CA).
Mussa, M. and S. Rosen, 1978, Monopoly and product quality, Journal of Economic Theory 18, 301–317.
National Telecommunications and Information Administration, 1988, Video program distribution and cable television: Current policy issues and recommendations, NTIA report no. 83-233, June (National Telecommunications and Information Administration, Washington, DC).
Perry, M.K., 1989, Vertical integration: Determinants and effects, In: R. Schmalensee and R. Willig, eds., Handbook of industrial organization (North-Holland, Amsterdam) 183–258.
Robichaux, M., 1995, Cable TV consolidates rapidly in the hands of a few, Wall Street Journal, Feb. 8, B4.
Scherer, F.M. and D. Ross, 1990, Industrial market structure and economic performance, 3rd Ed. (Houghton Mifflin, Boston, MA).
Spence, A.M., 1976, Product selection, fixed costs, and monopolistic competition, Review of Economic Studies 43, 217–235.
Tirole, J., 1988, The theory of industrial organization (MIT Press, Cambridge, MA).
United States Department of Justice and Federal Trade Commission, 1992, Horizontal merger guidelines (United States Department of Justice and Federal Trade Commission, Washington, DC).
Variety (weekly) (Cahners Publishing Co., Newton, MA)
Waterman, D. and A.A. Weiss, 1996, Vertical integration in cable television (MIT Press, Cambridge, MA; AEI Press, Washington, DC).
White, L.J., 1977, Market structure and product varieties, American Economic Review 67, 179–182.

Among other media industries to which the free rider model may apply, I focus on theatrical motion pictures. The model suggests motives behind the extensive horizontal and vertical integration and apparent cartel behavior in this industry prior to *U.S. v. Paramount Pictures, Inc.* et al. (1948).[7] In *Paramount*, the U.S. government achieved a major antitrust victory over eight motion picture distributors, five of which were integrated with movie theater chains. These integrated theater chains accounted for 70% of all 'first run' box office receipts at the time, and were heavily concentrated within local market areas; in 34 of the 85 largest U.S. cities, one chain controlled 75% or more of first run capacity; one chain controlled over 50% of capacity in 63 of the markets, and in these and other cities, 'pooling agreements' among theater owners limited competition (Loew's Exhibit L-13).

The government's case, which was basically accepted by the Supreme Court, was that the integrated distributors operated as a cartel in order to exchange access to each other's controlled theater markets, to the exclusion of independently owned distributors and theaters. The Court decision mandated complete vertical disintegration and extensive horizontal divestiture by the theater chains. The free rider model suggests that theater chain formation may have been driven by the incentive to exert monopsony power. The vertical integration (which mostly followed the theater chain formation in time) and eventual formation of the *Paramount* cartel may have been to limit opportunistic price setting behavior by the theater coalitions.[8] While other factors are clearly relevant, the model's welfare results suggest that the vertical integration and apparent cartel behavior among integrated motion picture firms may have served a pro-competitive function.

With minor tinkering, the free rider model can be applied to developing market structure and recent antitrust controversy in the pharmaceuticals industry. Patent drug manufacturers must recapture large R&D expenditures by selling their products to retail pharmacies, hospitals, and Health Maintenance Organizations (HMOs) at prices well above their marginal costs of production and distribution, at least on average (Caves et al., 1991). In this respect, cost structure of pharmaceuticals manufacturing and distribution is analogous to that of media product creation and distribution.

The free rider model suggests that recent growth of hospital chains, HMOs, and retail pharmacy chains may be motivated or encouraged by the benefits of 'free riding' in the wholesale purchase of patent drugs. Although the basis for monopsony bargaining power by these downstream drug buyers is more complicated than the simple accumulation of geographic local market territories, the underlying principles are analogous.[9]

[7] United States v. Paramount Pictures, et al., 1948, U.S. 334 US 1.

[8] Hampton (1970) and Lewis (1933) offer early histories of market structure formation in the motion picture industry.

[9] See, for example, the Frech (1978) analysis of monopsony power in health insurance markets.

In a recent and highly publicized class action suit, independently owned pharmacies successfully sued 22 drug manufacturers for discriminatory pricing (and price fixing) because of large discounts given by the manufacturers to HMOs, hospital chains, and mail order pharmacies (France, 1996). A Federal Trade Commission investigation into these alleged discriminatory pricing practices has followed. The alleged price differentials are analogous to the differentials between input prices paid by larger v. smaller programming buyers in the cable television case. To the extent that the free rider model applies, it suggests that powerful drug buyers may exert negative effects on the variety and quality of pharmaceutical products by reducing aggregate R&D expenditures. Such reductions may cause consumer welfare to fall. The ambiguity of the present model's welfare results, however, reminds us of the difficultly in making confident efficiency judgements where questions of product variety are concerned.

In conclusion, consider explicitly the distinction made in this paper between the *incentive* and the *ability* to exercise monopsony power at the local level. A recent landmark antitrust case also in the movie industry, *U.S. v. Syufy Enterprises and Raymond Syufy* (1990),[10] shows how a failure to recognize the significance of national market shares in determining that ability may lead to an *overestimate* of monopsony power.

In 1981, Syufy entered the first run Las Vegas movie theater market by building a new theater complex. Syufy then proceeded to buy out each of his three main competitors to obtain by 1984 a virtual monopoly of the first run theater market in Las Vegas. In 1985, the Justice Dept sued Syufy under the Sherman Act.

The government's case was not based on monopolization of the consumer market. In fact, the government admitted that Las Vegas ticket and concession prices were no higher than in comparable cities having competitive theaters. Rather, the government based its case on Syufy's alleged monopsonization against his Hollywood suppliers within the city of Las Vegas. (Six or seven firms controlled the national film distribution market). The government lost in the District Court, and that decision was then upheld in a notorious Appeals Court decision in which Judge Kozinski humiliated the government by incorporating the titles of over 200 classic movies into the written opinion.

The key premise of the Appeals Court decision was that although Syufy may have acquired a virtual monopoly of first run theater seats in Las Vegas, entry into theater operation was not difficult. Entry did in fact occur during the trial and appeal period, reportedly reducing Syufy's market share from 93% in 1984 to 75% in 1988. The decision also reported at some length that the Hollywood distributors consistently testified at trial that Syufy did not receive input terms any more favorable than those paid by competing theater operators in other cities, and that the distributors were satisfied with Syufy's terms.

The free rider model suggests a different interpretation of why the government's

[10] United States v. Syufy Enterprises and Raymond Syufy, 9th Circuit, 1990, 903 F.2d 659.