NBER WORKING PAPER SERIES

AFFILIATION, INTEGRATION, AND INFORMATION: OWNERSHIP INCENTIVES AND INDUSTRY STRUCTURE

Thomas N. Hubbard

Working Paper 8300
http://www.nber.org/papers/w8300

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2001

This research benefitted from conversations with George Baker, Luis Garicano, and Bill Vogt. I am grateful for the assistance of Ryan Taliaferro. All remaining errors are mine. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

Affiliation, Integration, and Information: Ownership Incentives and Industry Structure
Thomas N. Hubbard
NBER Working Paper No. 8300
May 2001
JEL No. L11, L22

ABSTRACT

This paper presents theory and evidence on horizontal industry structure, focusing on situations where plant-level scale economies are small and market power is not an issue. At issue is the question: what makes industries necessarily fragmented? The theoretical model distinguishes between the structure of brands and firms in an industry by examining trade-offs associated with affiliation and integration, and how they are affected by the contracting environment. I show how contractual incompleteness can lead industries to be necessarily fragmented. I also show that improvements in the contracting environment will tend to lead to a greater concentration of brands, but whether they lead industries to be more or less concentrated depends on what becomes contractible. I then discuss the propositions generated by the model through a series of case study examples.

Thomas N. Hubbard
University of Chicago
Graduate School of Business
1101 East 58th Street
Chicago, IL 60637
and NBER
thomas.hubbard@gsb.uchicago.edu

I. Introduction

Two core theoretical literatures underlie much of the field of industrial organization: theories of industry structure and theories of the firm. Although the issues these literatures emphasize are very different, they actually seek to explain the same thing. Theories of industry structure are, in a way, also theories of firms' boundaries. Likewise, theories of firms' boundaries can also be read as theories of industry structure.

These literatures have different strengths and weaknesses. The industry structure literature provides several good explanations for why industries might be necessarily concentrated. The traditional literature (Bain (1956), Scherer (1980)) emphasizes factors such as scale economies, market size, and entry barriers; the newer literature adds concepts such as heterogeneity in tastes, the strength of price competition, and consumers' sensitivity to sunk expenditures such as advertising (Spence (1976), Dixit and Stiglitz (1977), Salop (1979), Sutton (1991)). As I describe in more detail below, this literature provides less convincing explanations for why some industries might be necessarily fragmented. Understanding fragmentation is important in part because most industries are, in fact, fragmented. For example, single-establishment firms make up more than half of revenue in the service sector of the U.S. economy, and the average firm in the service sector consists of 1.1 establishments.[1] Recent theories of the firm such as Grossman and Hart (1986) explicitly address the costs and benefits of integration and provide explanations of fragmentation – why establishments might be necessarily independently-owned. These and other theories also suggest how integration-related trade-offs might be related to the contracting environment. But the existing models in this literature commonly assume gains from trade between establishment managers, and thus do not examine the interplay between affiliation-related and integration-related trade-offs.

This paper presents theory and evidence on horizontal industry structure, focusing on situations where plant-level scale economies are small and market power is not an issue.

I first present a canonical model that uses Grossman and Hart (1986) as a foundation, but relaxes the assumption that there are gains from trade between managers and extends the framework

---

[1]Bureau of the Census (1995). Many retail sectors are also fragmented. The average retail firm consists of 1.4 establishments, and only 65% of sales are in multiestablishment firms, despite the presence of several large, highly-concentrated segments such as discount retailing.

by allowing managerial effort to be contractible to varying degrees. In this model, a manager develops a business format that may be valuable if used at establishments other than that at which he works. There are two issues: whether the business format should be used at other establishments, and whether the manager who developed the business format should own the establishments at which it is used. The model distinguishes between the common use of business formats (affiliation) and common ownership (integration), and therefore conceptually distinguishes the concentration of brands and firms in an industry. The model's output is a series of comparative static predictions that relate the concentration of brands and firms to characteristics of establishment managers' jobs and the richness of the contracting environment. The latter is particularly important for applied research because variation in the contracting environment is sometimes observable to researchers – for example, when it corresponds to technological adoption or changes in the legal environment (Baker and Hubbard (2001)). A central theme is that improvements in the contracting environment tend to lead to a greater concentration of brands, but whether they lead industries to be more or less concentrated depends on what becomes contractible.

I then discuss the predictions of the model through a series of extended case study examples. Some of these examples come from existing quantitative research; others are newer and less quantitative. These examples show how the various trade-offs in the model are played out in the real world. Future research will test the model's predictions using systematic data.

The rest of the paper is organized as follows. Section 2 briefly motivates the analysis by discussing the shortcomings of the existing industry structure literature. Section 3 presents the model. I first discuss a version where, as in the "property rights" literature, agents' actions are completely non-contractible. I then present an extension that allows for varying degrees of contractibility; this extension is important because it generates comparative statics with respect to changes in the informational environment. Section 4 discusses the model and presents the case study evidence. Section 5 concludes.

II. Two Usual Suspects

Scale economies and the value of brands play important roles in the existing literature on industry structure, which focuses on the question: what factors lead industries to be concentrated? These factors, however, do not by themselves determine whether industries are fragmented.

While the presence of large scale economies implies that industries will be concentrated, the

converse does not hold: the absence of scale economies does not imply that industries will be necessarily fragmented. Standard replication arguments imply that such industries could either be concentrated or fragmented. Most existing theories of industry structure have neoclassical theories of the firm at their core, and the weaknesses of the neoclassical theories are exposed when trying to explain fragmentation.

To see this, consider an industry where the production technology does not imply significant scale economies – say, dry cleaning. Scale economies are very weak at the establishment level in dry cleaning. There is little or no cost disadvantage from operating a small dry cleaning outlet rather than a large one. This partly explains why most dry cleaning outlets are small. Furthermore, it is unlikely that there are scale economies from operating multiple outlets rather than single ones.[2] However, the absence of such economies does not explain why the dry cleaning business is so fragmented. While there may be no cost advantage from operating multiple outlets, this argument does not imply a disadvantage – the absence of scale economies does not imply that there exist scale diseconomies. Absent other factors, the industry could consist either of a few large chains of commonly-owned outlets or of many independent outlets. Since one objective of the analysis is to identify forces that cause fragmentation, scale economies will take a back seat in the analysis.

The second usual suspect is the value of brands. The main point here is that consumers' brand- or advertising-sensitivity may lead industries to be concentrated in terms of brands, but brand concentration does not necessarily imply firm concentration because brands can be produced by multiple firms. In this light, Sutton's (1991) seminal work on industry structure speaks at least as much to the concentration of brands, or business formats, in an industry as to the concentration of firms.

This point is perhaps clearest in the context of branded chains. Starbucks and Ace Hardware are both chains: each has many affiliated establishments that share the same brand name and basic business format. Starbucks and Ace differ in an important way, however. Almost all Starbucks outlets are owned by Starbucks Corporation.[3] Nearly all Ace outlets are independently-owned.

---

[2]Scale economies may be greater upstream, for example from operating a dry cleaning plant that serves multiple retail outlets.

[3]One important exception is in airports, where existing agreements between concessionaires and airport authorities make it hard for Starbucks to own these outlets.

Affiliation is valuable both in coffee and hardware retailing, but this fact alone does not necessarily imply that establishments that share business formats or brand names will be commonly-owned – and thus whether these industries will consist of few or many firms.

Reflecting on the hardware example for a bit is worthwhile, because the organizational alternatives in the model below are familiar in this context. Some hardware sellers are completely independent; they are owned by their manager and are not affiliated with other establishments. Some, like Ace affiliates, are part of independently-owned chains; they are owned by their manager and share a brand name with other outlets.[4] Some, like Home Depot stores, are part of commonly-owned chains. They are affiliated with other outlets but are not owned by their manager.

The fact that all three organizational forms appear in hardware retailing is somewhat unusual, but the point that three broad alternatives exist for an establishment is a general one. In general, establishments may or may not share a business format and those that share formats may or may not be owned by the same party. Note that I am using the phrase "business format" broadly, to refer to many potential sources of affiliation in addition to business practices, brand names, and trademarks. In some manufacturing contexts a business format may center around a proprietary technology; affiliations occur when multiple establishments use this technology, and therefore may be enabled by licensing arrangements. In certain retailing contexts, a business format may be a set of purchasing or advertising arrangements, and affiliations may provide for collective purchasing or advertising.

The next section presents a model for understanding when and why establishments have common formats and/or ownership.[5] Industries in which factors strongly favor independent ownership of establishments are necessarily fragmented.

III. The Model

Suppose an industry consists of a large number of establishments, each of which is run by an individual manager. Assume that establishments are price-takers in input and output markets, so there is no pricing advantage from merging. Assume also that there exist no production scale economies or diseconomies from operating multiple establishments, so there is no inherent cost

[4]In other contexts such as fast food, franchise outlets fall in this category when the franchisee owns the establishment. I discuss this further below. See also Maness (1996).

[5]This model is similar to that in Baker and Hubbard (2000), and draws heavily from Grossman and Hart (1986).

advantage or disadvantage from size.

Denote two of these establishments as A1 and A2; these are run by managers Victor and Paul, respectively. Suppose Victor develops a business format that, all else equal, raises the value of any individual establishment by K. He puts this in place at his establishment, A1.

The model investigates two questions. First, under what conditions should Victor's business format be used at A2 as well? That is, should A2 be part of a chain ("Victor's") or not? Second, who should own A2: Victor or Paul? In particular, if A1 and A2 are part of a chain, should this chain consist of independently- or commonly-owned establishments?

A2's Value and Paul's Decisions

I first specify A2's value under different business formats. Let V be A2's value if Victor's business format is used. Specify V as:

$$V = K + \sum_{i=1}^{g_1} e_1^i$$

Let P be A2's value if Paul's format is used. Specify P as:

$$P = \sum_{i=1}^{g_2} e_2^i + \sum_{i=1}^{g_1} e_1^i$$

Paul thus makes two classes of effort decisions. These are represented by vectors $e_1 = \{e_1^i\}$ and $e_2 = \{e_2^i\}$, which have dimension g_1 and g_2, respectively.

e_1 consists of actions that increase A2's value, regardless of whose business format is used. (Note that e_1 enters the expressions for V and P in the same way.) I will refer to activities in e_1 as "establishment management," or simply "management." e_1 could include actions that enhance the establishment's reputation with its customers – a reputation for a clean store or friendly service might increase a store's value regardless of its business format. This interpretation would be particularly salient in service or retail contexts. This is not the only possible interpretation, however. e_1 could also include actions that maintain the value of physical plant, or cost-reducing innovations that have industry-wide value. For simplicity, I assume that the individual actions e_1^i have equal marginal productivity, and normalize this marginal productivity to one. Each of these actions represent a distinct opportunity to affect A2's value.

g_1 is the dimension of e_1. The higher g_1, the more opportunities Paul has to affect A2's value. Following from above, g_1 will tend to be high when A2's value is reputation-sensitive, when managerial actions are important for maintaining physical assets' value, or when managers are important sources of industry-wide innovations.

e_2 consists of a vector of actions that increase A2's value if Paul's business format is used, but not if Victor's is used. (Note that e_2 enters the expression for P but not V.) e_2 therefore includes activities such as generating alternative business formats for A2, lining up clients for such formats, and producing process or product innovations that are incompatible with Victor's format. These are activities that would be considered *entrepreneurial* conditional on Paul's format being used, but *opportunistic* conditional on Victor's. This captures the idea that entrepreneurial and opportunistic activities are often differentiated by their context rather than their substance.[6] I will henceforth refer to e_2 as Paul's "entrepreneurial activities" with the understanding that such activities may or may not be desirable.

g_2 is the dimension of e_2. If g_2 is low, Paul has few opportunities to engage in entrepreneurial activities. This would be the case when managers' efforts toward developing business formats are unlikely to be productive. This may be true if A2 were a large discount retailing outlet (a Wal-Mart), for example. If g_2 is high, Paul has many opportunities to do things that increase A2's value but are incompatible with other establishments' format. This might be the case if A2 were a high-end restaurant, where a reputation for good "specials" would increase demand. As above, I assume that the individual actions e_2^i have equal, unitary marginal productivity.

I assume that effort is costly to Paul, and these costs are represented by the equation:

$$C = \sum_{i=1}^{g_1} \tfrac{1}{2}(e_1^i)^2 + \sum_{i=1}^{g_2} \tfrac{1}{2}(e_2^i)^2$$

Costs are thus convex in e_1^i and e_2^i, and e_1^i and e_2^i are neither complements nor substitutes in Paul's effort supply function. This assumption makes choices of e_1^i and e_2^i independent of each other. This corresponds to a situation where Paul faces a series of independent opportunities to affect A2's value. Independence means that, unlike in Holmstrom and Milgrom (1991, 1994), managers do not face meaningful effort allocation decisions. This will be useful later in the paper, because it will

[6]This is noted by Holmstrom (1999).

mean that contracting on one dimension of effort will not distort effort with respect to other dimensions.

This cost specification means that increasing g_1 or g_2 decreases the marginal cost of managerial effort summed over activities. It therefore will lead to more managerial effort in some of the cases examined below, because the marginal benefit of managerial effort summed over activities is independent of g_1 and g_2. This particular set of assumptions is for modeling convenience, and is not crucial to the results. One can produce the same comparative statics in an algebraically-messier model where the marginal benefit of effort (summed over activities) increases in g_1 and g_2 but the marginal cost of effort is constant in these variables. One can therefore interpret variation in g_1 or g_2 either as differences in marginal productivities or in the number of managerial opportunities to affect assets' value.

Timing and Assumptions

The timing and assumptions are similar to Grossman and Hart (1986) and related papers. There are three stages. First, Victor and Paul negotiate over which format should be used for A2 and who should own it. I assume that they make efficient decisions with respect to affiliation and ownership. These decisions are renegotiable, and both parties realize this at the start. Second, Paul chooses e_1 and e_2 to maximize his own utility, conditional on the format and ownership decisions made in the first stage. Third, Victor and Paul observe V and P and negotiate over how the returns from A2 (either V or P, depending on the format chosen) should be split. They may also negotiate over format and ownership at this point, although in the model format and ownership changes at this stage are not equilibrium outcomes.

I assume that asset ownership allows individuals to decide how assets are used in circumstances not covered by existing contracts. Further, I assume for simplicity that assets' use is completely non-contractible, and therefore that all control rights are residual. An important implication of these assumptions is that Paul has the option of unilaterally operating A2 under his own format if and only if he owns A2. I also assume that V, P, and managerial effort choices $\{e_1^i\}$ and $\{e_2^i\}$ are non-contractible. Later I relax the assumption on the non-contractibility of effort.

Finally, I assume that the outcome of bargaining in the third stage is the Nash bargaining solution. Thus, how Victor and Paul divide surplus depends on their outside options. I normalize Victor's outside option to zero, regardless of whether he owns A2. I assume that Paul's outside

option differs depending on whether he owns A2. If he owns A2, his outside option is to operate A2 under his own business format. Therefore, his outside option is equal to P. If Paul does not own A2, he does not have the option of operating A2 under his own business format. I will assume that his outside option is some constant (his wage in another job), which I will call W.

I solve the model by first examining Paul's decision under each organizational option, then comparing the surplus created under each option.

Case 1: Paul's format is used, Paul owns A2. (A2 is completely independent.)

If Paul owns A2 and uses his own format, there is no trade and thus no bargaining between Victor and Paul. Paul appropriates the full value of his activities with respect to A2 if A2 is operated as a completely independent outlet. He therefore chooses e_1 and e_2 to maximize P – C, the value of his establishment net of effort costs. This equals:

$$P - C = \sum_{i=1}^{g_1} e_1^i + \sum_{i=1}^{g_2} e_2^i - \sum_{i=1}^{g_1} \tfrac{1}{2}(e_1^i)^2 - \sum_{i=1}^{g_2} \tfrac{1}{2}(e_2^i)^2$$

Paul therefore chooses $e_1^i = 1$, $e_2^i = 1$. Plugging these choices into the expressions for P and C, surplus equals $(g_1 + g_2)/2$.

Case 2: Paul's format is used, Victor owns A2.

Nash bargaining implies that Paul receives (P-W)/2 + W = (P+W)/2. He thus maximizes:

$$(P + W)/2 - C = \tfrac{1}{2}(\sum_{i=1}^{g_1} e_1^i + \sum_{i=1}^{g_2} e_2^i + W) - \sum_{i=1}^{g_1} \tfrac{1}{2}(e_1^i)^2 - \sum_{i=1}^{g_2} \tfrac{1}{2}(e_2^i)^2$$

and chooses $e_1^i = 1/2$, $e_2^i = 1/2$. This yields $3(g_1 + g_2)/8$, which is always less than the surplus in case 1. This case is therefore never relevant: if Paul's format is used, he should own A2. Giving Victor ownership provides no incentive benefits, and discourages surplus-increasing effort from Paul.

Case 3: Victor's format is used, Paul owns A2. (A2 is part of a chain of independently-owned outlets.)

Paul's share of the surplus is (V-P)/2 + P = (V+P)/2. He chooses e_1 and e_2 to maximize

(V+P)/2 – C, which is equal to:

$$\tfrac{1}{2}(K + 2\sum_{i=1}^{g_1} e_1^i + \sum_{i=1}^{g_2} e_2^i) - \sum_{i=1}^{g_1} \tfrac{1}{2}(e_1^i)^2 - \sum_{i=1}^{g_2} \tfrac{1}{2}(e_2^i)^2$$

He therefore chooses $e_1^i = 1$, $e_2^i = 1/2$. The surplus from this case is K + ½ g_1 – 1/8 g_2.

Case 4: Victor's format is used, Victor owns A2. (A2 is part of a commonly-owned chain.)

If Paul does not own A2, he maximizes (V+W)/2 – C, which is equal to:

$$\tfrac{1}{2}(K + \sum_{i=1}^{g_1} e_1^i + W) - \sum_{i=1}^{g_1} \tfrac{1}{2}(e_1^i)^2 - \sum_{i=1}^{g_2} \tfrac{1}{2}(e_2^i)^2$$

He therefore chooses $e_1^i = 1/2$, $e_2^i = 0$; surplus equals K + 3/8 g_1. Comparing cases 3 and 4, e_1 and e_2 are higher when Paul owns A2 than when he does not. This is because he is better able to appropriate the returns from his efforts when he owns A2.

Table 1
Effort and Surplus – All Effort Non-Contractible

	Effort	Surplus
Completely independent	$e_1^i = 1$, $e_2^i = 1$	½ g_1 + ½ g_2
Chain, independently-owned	$e_1^i = 1$, $e_2^i = 1/2$	K + ½ g_1 – 1/8 g_2
Chain, commonly owned	$e_1^i = 1/2$, $e_2^i = 0$	K + 3/8 g_1

Table 1 summarizes the results to here and illuminates the trade-offs with respect to the three relevant organizational options.

If A2 is kept completely independent, Paul will choose high levels of both e_1 and e_2, and this is good. However, the parties forego the opportunity to use Victor's business format, which may be valuable.

If A2 is part of a chain of independent outlets, Paul has strong incentives with respect to e_1 – management activities that increase A2's general value – and this is good. However, Paul also has

fairly strong incentives to choose high levels of e_2. This is bad, given that Victor's format is being used – Paul's entrepreneurial activities are privately valuable because they help Paul appropriate a larger share of the returns from A2. However, they do not actually increase A2's value because they are never optimal to actually implement.

If A2 is part of a commonly-owned chain, Paul has weak incentives with respect to e_1, and this is bad. But he also has weak incentives with respect to e_2, which is good in light of the discussion above. Given that Victor's format is being used, it is better that Paul does not spend time on entrepreneurial activities.

This highlights the benefits and drawbacks of allowing Paul to own A2 under Victor's format. Ownership creates both good and bad incentives. It strengthens incentives both with respect to increasing the value of the asset (e_1) and to investments in bargaining positions (e_2). The former is always good; the latter is good only when Paul's format is used. This characterization of organizational trade-offs is consistent with managerial sentiment that while owning affiliated enterprises may weaken managerial incentives, it also allows for more control over these enterprises.

Results

Figure 1 summarizes the results. The three regions in this figure correspond to situations where each of the three organizational choices are optimal. The optimal format and ownership combination depends on two ratios: K/g_2 and g_1/g_2.

Proposition 1: Chains are optimal when K/g_2 is high, but not otherwise.

The intuition behind this is simple. K/g_2 reflects whether the manager of an establishment has a comparative advantage in developing the establishment's business format. If g_2 is low relative to K, the manager does not have a comparative advantage – so becoming part of a chain is optimal. If g_2 is high relative to K, the manager has a comparative advantage – so being independent is optimal.

It is useful to note at this point why the line dividing "chain, commonly-owned" and "completely independent" is upward-sloping. The intuition is this. If $g_1 = 0$, giving Paul management incentives offers no benefits, so if $g_2 > 0$, the relevant chain option is "commonly-

owned chain." The decision to operate A2 as part of a chain or not is made purely on the basis of whether Victor or Paul has a comparative advantage in developing the outlet's business format; that is, whether K/g_2 is greater than 1/2. Increasing g_1 slightly introduces an incentive cost to having A2 part of a chain; the slope of the dividing line reflects this incentive cost. If Victor's comparative advantage in developing the outlet's business format is very small, the comparative advantage will not outweigh the incentive cost and the outlet will remain completely independent.

Proposition 2: Establishments should be owned by their managers if g_1/g_2 is high, even when chains are optimal.

When g_1/g_2 is high, the benefits of having the manager own the establishment are high relative to the drawbacks. The benefit of managerial ownership is that it strengthens managers' incentives to do things that increase the establishment's value. If g_1 is high, managers have many opportunities to increase the establishment's general value so the benefits of managerial ownership are high. The drawback associated with chains of independently-owned outlets is that ownership gives managers incentives to engage in unwanted entrepreneurial activities. But when g_2 is low, managers have few opportunities to do so. The drawback to managerial ownership is therefore low.

Proposition 3: Industries are necessarily fragmented unless K/g_2 is high and g_1/g_2 is low, the upper-left region of Figure 1.

The intuition behind this follows from propositions 1 and 2, which answer the question: why are commonly-owned chains not optimal in the other regions? When K/g_2 is low it is better to have the manager develop his or her own business format – so chains are not optimal. When g_1/g_2 is high, the benefits of having the manager own the establishment are high relative to the drawbacks, even when the store is part of a chain.

One way to think about the upper left region is that it is where ownership incentives are unimportant for motivating managers. Motivating the manager to be entrepreneurial is unimportant, because using an existing business format is optimal (K is high). Furthermore, motivating establishment management is relatively unimportant, because managers have few opportunities to

do things that increase the establishment's general value (g_1 is relatively low).

Managerial incentives are unimportant in most theories of industry structure. Such theories are implicitly considering situations in this region of Figure 1.

Ownership incentives are more important in the other regions. If the manager owns the establishment, he or she has a greater incentive both to increase the value of the establishment and be entrepreneurial. If either of these classes of activities are valuable and managerial effort is non-contractible, it may be best for the manager to own the establishment. In such circumstances, industries will necessarily be fragmented (although one may observe independently-owned firms that share business formats).

The non-contractibility assumption is important in the model above. I next extend the basic model by allowing managerial effort to be contractible to varying degrees. This will generate a set of comparative statics that relates variation in industrial organization to variation in the contracting environment. This is important because, as emphasized in Baker and Hubbard (2001), variation in the informational environment is sometimes more observable to applied researchers than variation in K, g_1, or g_2.

Contractibility and Industrial Organization

Partition the g_1 and g_2 dimensions of managerial effort so that n_1 and n_2 of them are non-contractible and g_1-n_1 and g_2-n_2 are contractible, respectively. Assume that effort decisions that are contractible are perfectly contractible, so that one can elicit first-best effort levels costlessly. V and P then become:

$$V = K + \sum_{i=1}^{n_1} e_1^i + \sum_{i=n_1+1}^{g_1} e_1^i$$

$$P = \sum_{i=1}^{n_2} e_2^i + \sum_{i=n_2+1}^{g_2} e_2^i + \sum_{i=1}^{n_1} e_1^i + \sum_{i=n_1+1}^{g_1} e_1^i$$

Table 2 summarizes the results from solving the model under this new set of assumptions.

Under the "completely independent" option, nothing changes. Paul chooses first best levels of e_1 and e_2, regardless of the contracting environment. Effort and surplus are the same as before.

Under the "independent chain" option, the contractible decisions are set at first-best given the format, which is $e_1^i = 1$, $e_2^i = 0$. Non-contractible decisions are as before, $e_1^i = 1$, $e_2^i = 1/2$. Surplus is $K + g_1/2 - \mu_2 g_2/8$, where $\mu_2 = n_2/g_2$ is the fraction of entrepreneurial decisions that are non-contractible.

If A2 operates as part of a commonly-owned chain, the contractible decisions are once again set at $e_1^i = 1$, $e_2^i = 0$. The non-contractible decisions are $e_1^i = 1$, $e_2^i = 1/2$, and total value is $K + g_1(1/2 - \mu_1/8)$, where $\mu_1 = n_1/g_1$ is the fraction of non-entrepreneurial decisions that are non-contractible.

Table 2
Effort and Surplus – Effort Partly Contractible

	Contractible Effort	Non-Contractible Effort	Surplus
Completely independent	$e_1^i = 1$, $e_2^i = 1$	$e_1^i = 1$, $e_2^i = 1$	$\frac{1}{2} g_1 + \frac{1}{2} g_2$
Chain, independently-owned	$e_1^i = 1$, $e_2^i = 0$	$e_1^i = 1$, $e_2^i = 1/2$	$K + \frac{1}{2} g_1 - \mu_2 g_2/8$
Chain, commonly owned	$e_1^i = 1$, $e_2^i = 0$	$e_1^i = 1/2$, $e_2^i = 0$	$K + (1/2 - \mu_1/8) g_1$

Figures 2 and 3 depict how the organizational options depend on μ_1 and μ_2, the extent to which managerial decisions are contractible. I next discuss a series of propositions using these figures.

Figure 2 shows the effect of decreasing μ_1, the fraction of managerial effort that is non-contractible. Decreasing μ_1 shifts the vertical line separating the two chain regions to the right and flattens the diagonal that separates commonly-owned chain from completely independent.

Proposition 4: Increases in the contractibility of establishment management should shift outlets from completely independent and independently-owned chains to commonly-owned chains. Thus, such increases should make industries more concentrated and lead to more chains.

Decreases in μ_1 affect two margins. They shift outlets that are part of chains from independently-owned to commonly-owned by eliminating an advantage of independent ownership. If establishment management becomes contractible, then one can elicit first-best effort levels with a contract and discourage unwanted entrepreneurial effort by denying the manager ownership.

Decreasing μ_1 also shifts completely independent outlets to commonly-owned chains – some outlets change both ownership and affiliation. The intuition for this follows the discussion under Proposition 1. Starting from $g_1 = 0$, increasing g_1 introduces an incentive cost to operating A2 as part of a chain; thus, some outlets will remain independent if the outlet manager's comparative disadvantage in developing the outlet's format is sufficiently small. But if establishment management activities are mostly contractible, the incentive cost will be small. Decreasing μ_1 means that whether outlets are part of a chain will be based more on comparative advantage issues, and less on incentive issues.

It is important for empirical testing to recognize two additional results from this diagram – where the model predicts that changing μ_1 should have no effect.

Proposition 5: Changes in μ_1 should not affect industrial organization if K/g_2 is low.

Proposition 6: Changes in μ_1 should not lead outlets to move from completely independent to independently-owned chain, or vice-versa.

Note that the organizational effect of decreasing μ_1 is exactly the same as decreasing g_1: increasing the contractibility of establishment management has the same impact as decreasing the marginal productivity of such effort. This formalizes the argument Baker and Hubbard (2000) use to generate comparative statics with respect to on-board computer adoption in trucking. Below I show that this equivalence result does not hold for μ_2 and g_2.

Figure 3 shows the effect of decreasing μ_2, the fraction of entrepreneurial effort that is non-contractible. Decreasing μ_2 shifts the borders of "chain, independently-owned" down and to the left.

Proposition 7: Increases in the contractibility of entrepreneurial effort should shift outlets from commonly-owned chains and completely independent to independently-owned

chains. Thus, increases in the contractibility of such effort should make industries less concentrated but lead to more chains.

Making e_2 more contractible reduces the drawbacks associated with "independently-owned chain" by allowing parties to discourage unwanted entrepreneurial effort with contracts. Within a chain, this allows firms to achieve the benefits of giving managers' ownership incentives at a lower incentive cost. It will also lead to fewer independent outlets and thus more chains in general. If g_1/g_2 is high, the relevant chain option involves independent ownership. The incentive costs associated with chains diminish in this region as entrepreneurial effort becomes more contractible.

As above, there are two additional propositions related to where changes in μ_2 should have no effect.

Proposition 8: Changes in μ_2 should have no impact if K/g_2 is low.

Proposition 9: Changes in μ_2 should not lead outlets to move from completely independent to commonly-owned chain, or vice-versa.

As noted above, the comparative statics of g_2 are different than those of μ_2. (See Figure 4.) This is because changing g_2, Paul's entrepreneurial opportunities, has two effects. Increasing g_2 raises the incentive costs associated with the independently-owned chain option, thereby pushing outlets away from this option and toward the others. This effect is the same as when one increases μ_2. The other effect is different: increasing g_2 also raises the comparative advantage of using Paul's business format. This can shift outlets from commonly-owned chain to completely independent, something that the model predicts will not happen with a decrease in μ_2.

Finally, note that as either μ_1 or μ_2 approaches zero – that is, as either class of effort becomes completely contractible – the border separating the chain options from the completely independent option approaches the horizontal line $K/g_2 = 1/2$. Thus, contractual incompleteness leads to fewer chains than first-best, and therefore affects the concentration of brands as well as firms.

IV. Discussion and Examples

In the introduction, I reported that service industries tend to be very fragmented. This

framework suggests why: g_1 and μ_1 tend be high in service contexts. Managers tend to have many opportunities to increase establishments' value, and their effort with respect to these opportunities tends to be difficult to evaluate because it can be hard to find good performance measures. This tends to make it desirable for managers of service establishments to be owners, and sharply limits the concentration of service industries.

In an earlier era, manufacturing industries were extremely fragmented as well. Before the mid-1800s, there virtually no large firms in the U.S. No single establishment had more than 250 workers, and nearly all firms were single-establishment enterprises. By the end of the 19th century, however, there were many large multi-establishment firms. Chandler (1977) investigates the sudden growth of firms during this period, and attributes it to a set of complementary innovations. Some of these were technological improvements that increased the efficiency of mass production relative to craft production. Others were improvements in transportation – the railroad – that enabled firms to distribute goods over wider areas. Interestingly, Chandler also cites innovations in business practices – particularly the development of cost accounting – that, among other things, made evaluating managers of large, far-flung enterprises easier. Such improvements lowered μ_1, and thus played a role in the defragmentation of many manufacturing industries.

Below I provide several more recent and specific examples.

Banks

Banks supply many types of financial services, some of which involve managerial discretion. For example, consider the decision whether to grant a small business a loan. Bank managers use objective factors – such as the entrepreneur's credit history – in making such decisions. However, subjective factors can be informative as well; for example, a manager may be able to augment information from objective factors from a meeting with the entrepreneur, learning (sometimes difficult to quantify) local economic trends, and so on. It is difficult to use performance incentives to motivate managers to obtain information and make appropriate decisions on loan applications. For example, basing managers' bonuses on default rates would encourage them to only accept loans from "sure bets." Furthermore, whether entrepreneurs default on loans may only be known long after bank managers have left the bank for other positions. These and other factors impede the use of performance incentives.

Motivating bank managers to garner and process subjective information about borrowers'

prospects is important – that is, g_1 is high – when loans are to businesses whose prospects are difficult to quantify. Furthermore, μ_1 is high because it is difficult to motivate bank managers to do so contractually. One would therefore predict that manager-owned establishments should tend to handle a disproportionate share of small business loans. Recent research is consistent with this prediction. Brickley, Linck, and Smith (2000) find that ownership patterns in bank offices reflect the composition of local demand. In rural locations, where commercial loans tend to be to small businesses and farmers, bank offices tend to be owned by their managers. In cities, where commercial loans tend to be to large businesses, bank offices tend not to be owned by their managers – rather, they are part of a large chain of commonly-owned branches. Consistent with Figure 1, establishments tend to be independently owned in segments of banking where g_1 and μ_1 are high, but commonly owned in segments where g_1 is low.[7]

A similar logic may explain why lenders of venture capital, who operate in a segment where there are few objective measures of lenders' prospects and thus high returns to collecting and processing subjective information, are generally independently-owned enterprises.

Banking also provides a good example of why affiliation need not imply common ownership, and therefore why increases in the value of chain operation need not lead industries to become heavily concentrated. Affiliations among individual bank offices are valuable for many reasons. For example, they can facilitate check-clearing and other back-office processes and help pool loan portfolios, thereby lowering risk. Affiliations can also raise consumers' willingness to pay for the bank's services, for example, by allowing customers of one office to obtain cash from another office's ATM. K is thus fairly high in banking, and has increased with the emergence of ATMs. Nevertheless, a high and increasing K has not led to the extinction of independently-owned banks. Increases in K cause establishments to move up in Figure 1, thereby causing stand-alone offices to either become part of commonly owned chains (i.e., converted to being a branch of a large bank) or become part of a network of independently-owned banks. ATM networks and "bankers' banks" – institutions that provide transaction processing and loan pooling services to independently-owned

[7]Peterson and Rajan (2000) show that the average distance between small businesses and their lenders have declined over time, and attribute this to informational and communication improvements that have diminished local banks' comparative advantage in assessing credit risks. Similar themes appear in Lamoreaux's (1994) account of lending practices in 19th century New England.

member banks – allow independent banks to garner the benefits of affiliation in a way that does not involve common ownership.

Gas Stations

Shepard (1993) provides evidence from gasoline retailing. Gasoline stations are affiliated through the brand name of the gasoline they sell, and are either commonly-owned (usually by a refiner) or independently-owned (usually by their manager). Shepard finds that whether refiners own stations depends on the ancillary services they offer. Stations that have convenience stores tend to be commonly-owned; stations that have repair shops tend to be independently-owned. Shepard argues that this reflects differences in the contractibility of managerial effort between the ancillary services: it is easy to evaluate convenience store managers, thus motivating them with performance incentives works well. But it is hard to do so with auto repairers. In the current framework, μ_1 is higher for auto repairers than convenience store managers.

Shepard's results imply that as convenience stores replace repair shops as ancillary services, more stations should become part of commonly-owned chains and industry structure in gasoline retailing should become less fragmented. In fact, during the past twenty years, as convenience stores have supplanted repair shops at service stations, refiner ownership of gas stations has increased.

Trucking

Baker and Hubbard (2000) examine similar issues in the context of trucking. The question they examine is whether drivers own the trucks they operate. Affiliations arise across drivers and trucks through common dispatch, and are valuable when dispatchers have a comparative advantage relative to drivers in finding hauls for the truck. The advantage of using owner-operators is that drivers take better care of trucks when they own them. Driver ownership of trucks is costly within affiliations because ownership encourages them to engage in unwanted entrepreneurial activities. Identifying hauls other than the one their dispatcher wants them to accept can give them leverage in negotiating with dispatchers. Baker and Hubbard show that the introduction of on-board computers – technologies that let firms monitor how drivers operate trucks – has led to less driver ownership of trucks. Consistent with Figure 2, decreases in μ_1 have led affiliations to involve less asset ownership by their users, and thus have led the industry to become somewhat less fragmented.

Video Stores

The evolution of the video rental industry offers another nice example.[8] Video rental stores began to appear in the early 1980s in most parts of the U.S. At the time, nearly all were small, independently-owned outlets that had no chain affiliation. A few outlets were part of small chains such as Video Connection and Video Station, but even these were independently-owned franchises.[9] Since then, the organization of the video rental industry has become far less fragmented, especially during the mid-1990s. Between 1992 and 1997, single-establishment firms' share of storefronts fell from 69% to 54%, while the share of establishments part of firms with ten or more establishments increased from 13% to 34%.[10] Why are there commonly-owned chains now, but not earlier?

Video rental tastes differ sharply across local markets, even at the neighborhood level within metropolitan areas. It is important that stores are stocked with movies that appeal to local customers' tastes. Traditionally, store managers have had a comparative advantage in assessing local customers' tastes – for example, they receive feedback from customers when they take out and return movies. Managers' stocking decisions are difficult to evaluate using objective measures (how much of a store's current and future sales are due to good stocking decisions?); thus, performance incentives would generally not work very well in motivating managers to expend effort toward learning local tastes and making good inventory decisions. g_1 and μ_1 have traditionally been high, favoring independent ownership and leading to a fragmented industry.

During the mid-to-late 1980s, information technology in video stores became more sophisticated. Bar-coders and scanners that could capture transaction level data electronically began to diffuse in the industry. Capturing detailed point-of-sales data enabled firms to analyze store-level sales and inventory patterns centrally, and reduced store managers' comparative advantage in learning local tastes. Store-level stocking decisions now can be made just as well (or better) by a central office as by a store manager. In the model, moving local market analysis and purchasing decisions away from managers would lower g_1, and move outlets to the left in Figure 1; IT-related changes in managers' jobs should lead to more commonly-owned chains and a more concentrated industry.

In fact, accounts from the trade press suggest that the adoption of inventory tracking and

[8]This example evolved from discussions with Judy Chevalier.
[9]Paikert (1982).
[10]Bureau of the Census (1995b, 2000).

analysis systems and the emergence of large commonly-owned chains are related. The two chains that are cited by the trade press in the late-1980s as the most sophisticated users of inventory management technologies, Blockbuster and the now-defunct Erol's, were also among the first chains in which stores were not owned by their managers.[11] The large chains that have emerged since then, such as Hollywood Video, Video Update, and Movie Gallery all have sophisticated centralized inventory tracking systems, and operate chains of hundreds of commonly-owned outlets. These companies' securities filings commonly cite their use of point-of-sales technologies as critical to their business, claiming that they are crucial to store-level purchasing and inventory decisions.[12] Consistent with the theory, the diffusion of these technologies appears to have strongly contributed to the rapid emergence of commonly-owned chains and thus to the defragmentation of the industry.[13]

Uniformity and Industry Structure in Franchising[14]

Franchising began shortly before the end of the 19th century. Early franchises were for the rights for the sale or distribution of manufactured items such as sewing machines, automobiles, and gasoline. Franchise agreements generally provided franchisees the exclusive right to sell or distribute goods within well-defined territories. Territories were often very wide, sometimes encompassing entire states. These agreements allocated few ownership rights to franchisers. For example, franchisers generally had no control rights over any of the land or equipment used by the franchisee, and franchisees were usually free to develop their territories as they wished.

Franchising became much more common after World War II, as various firms began to franchise restaurant formats; examples include Dairy Queen, Big Boy, and of course, McDonald's. At first, these firms applied the same organizational form as their predecessors, but found that they had major problems motivating franchisees to conform to agreements regarding business practices. Although franchisers could stipulate operating procedures in great detail as part of franchise agreements, courts viewed franchisees as independent businessmen and were unwilling to enforce

[11]Keefe (1988), Chain Store Age Executive (1990), Chakravarty (1988), Wiener (1991).
[12]See, for example, Movie Gallery's 2000 10-K.
[13]Other factors may have increased K during this time, thereby encouraging the formation of chains. So the emergence of Blockbuster may reflect two factors: an increase in the effectiveness of advertising – which led to the formation of chains – and an increase in the contractibility of managers' decisions -- that led chains to be commonly owned.
[14]Much of this section draws from Love's (1995) outstanding account of McDonald's and other franchisers.

provisions that constrained how they ran their business. This was a particular problem at McDonald's, where cross-outlet uniformity in virtually all aspects of operations was considered an important source of the business format's value. Early franchisees departed from the format in various ways, including purchasing inputs from non-approved suppliers, adding menu items, customizing food for individual customers (e.g., grilling onions), and changing operating procedures to suit their tastes. These departures were a form of free-riding when they hurt other outlets within the chain.

McDonald's addressed this problem by departing from existing franchising practices in two ways. First, it retained the right to determine whether franchisees could open additional outlets, even when franchisees were granted exclusive territories. Although it could not easily terminate a non-conforming franchise through court action, it could deny its expansion. Second, after it had sold its first few franchises, it began to hold control rights over the land and buildings in which franchises operated. The following account explains the organizational logic:

> "Kroc and Sonnenborn [Kroc's partner at the time] believed that control of the real estate also gave McDonald's the type of control over the franchisee that it wanted but could not get from a franchise arrangement. Dozens of court cases have since defined the rights of the franchisee and the powers of the franchiser, but franchising was not well recognized in law in the 1950s. What was to prevent a franchisee from taking down the McDonald's sign, changing the restaurant name, and withholding his royalty fee? What power did McDonald's have to discipline the recalcitrants in its operator community who did not want to abide by the rules on menu or operating standards? In a battle with such malcontents, Sonneborn said, 'I never thought the franchise contract was worth the paper it was written on...'
>
> "Such was not the case with a lease. It was a time-honored legal document, and McDonald's quickly made compliance with its operating standards one of the requirements of the lease. 'We connected the lease to the franchise so that any violation of the franchise could create termination of the lease,' Sonneborn explained...
>
> "Sonneborn's thinking was extremely appealing to Ray Kroc in the mid-1950s when he was having such difficulty with...[early] franchisees, most of whom owned their units and leased the property from landowners... 'I have finally found the way that will put every single McDonald's we open under our complete control,' an excited Kroc reported...in early 1957, as he explained Sonneborn's plan. 'It [the lease] says that if at any time...that the

> operation does not conform in every way to the McDonald's standards of quality and service, this lease will be cancelled on thirty-day notice. Now we will have a club over them, and by God, there will be no more pampering or fiddling with them. We will do the ordering instead of going around and begging them to cooperate.' "[15]

McDonald's practice of retaining control rights over the establishment itself was quickly copied by many other franchised chains, including Southland's 7-11 stores, and more recently, Subway.

When uniformity is an important element of a business format, outlet managers have a stream of opportunities to depart from this, effectively creating a new business format based on the original. Thus, g_2 tends to be high. μ_2 tends to be high as well when courts tend not to enforce contractual provisions regarding managers' "entrepreneurial activities." In light of the model, it is thus not surprising that in circumstances like McDonald's where uniformity is important, the owner of the format also owns individual establishments' physical assets. It is also not surprising that manager ownership has been problematic when uniformity is valuable but difficult to motivate contractually.

In the several of the examples above, the difficulty of motivating e_1 contractually led industries to be necessarily fragmented. Technological change that mitigated this contracting problem led the trucking and video rental industries to become more concentrated. Here, the situation is reversed. Contractual difficulties led chains to be commonly-owned where uniformity is important, and thus made the industry more concentrated than it otherwise would be. It follows that contractual improvements in fast-food due to legal or technological changes in the contractibility of entrepreneurial activities would lead the industry to become more fragmented: managers would own their establishments more.

The general lesson is that contractual difficulties can led industries to be more or less fragmented, depending on the nature of the difficulties.

V. Conclusion

Different strains of the industrial organization literature examine the configuration of affiliations and ownership within industries. This paper presents a theory that examines these

[15]Love (1995), p. 156-7.

jointly, and demonstrates how affiliation and ownership patterns relate to the characteristics of managers' jobs and the contracting environment. I show how and why contractual incompleteness can lead industries to be necessarily fragmented. I also show that while improvements in the contracting environment tend generally to lead to more chains, they can lead industry structure to be more or less concentrated depending on what becomes more contractible. Finally, I present a series of case examples from historical accounts, the existing empirical literature, and the trade press that show how the relationships depicted in the model appear in the real world.

The theory and evidence in this paper are foundational. On the theoretical front, the canonical model can be extended in various ways. One can relax assumptions to allow multiple individuals' decisions to affect the gains from trade, permit individuals' effort decisions to be related across dimensions, or move away from single-manager establishments. This would allow one to analyze how organizational trade-offs involving appropriation, effort allocation, and delegation examined by other authors interact with those examined here. On the empirical front, the next steps are clear. The case narratives in this paper are consistent with several of the model's propositions. But they are more suggestive than conclusive. Additional research is needed to test the model's propositions more rigorously, using systematic data. Such evidence would help determine whether the basic trade-offs examined herein are systematically important, improve our understanding of fragmentation, and provide empirical links between two major strains of the industrial organization literature.

Figure 1
Business Formats and Ownership of Establishments

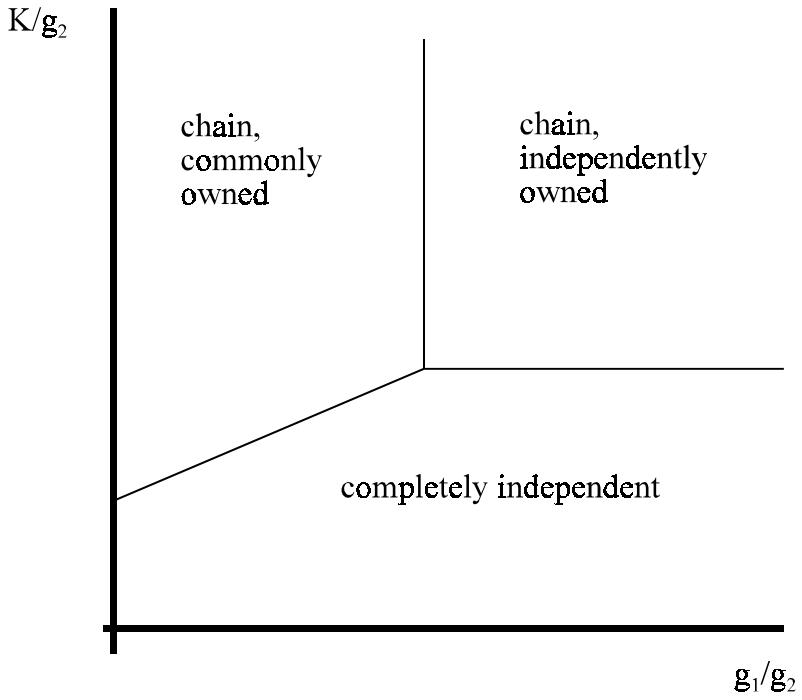


Figure 2
Increasing the Contractibility of e_1

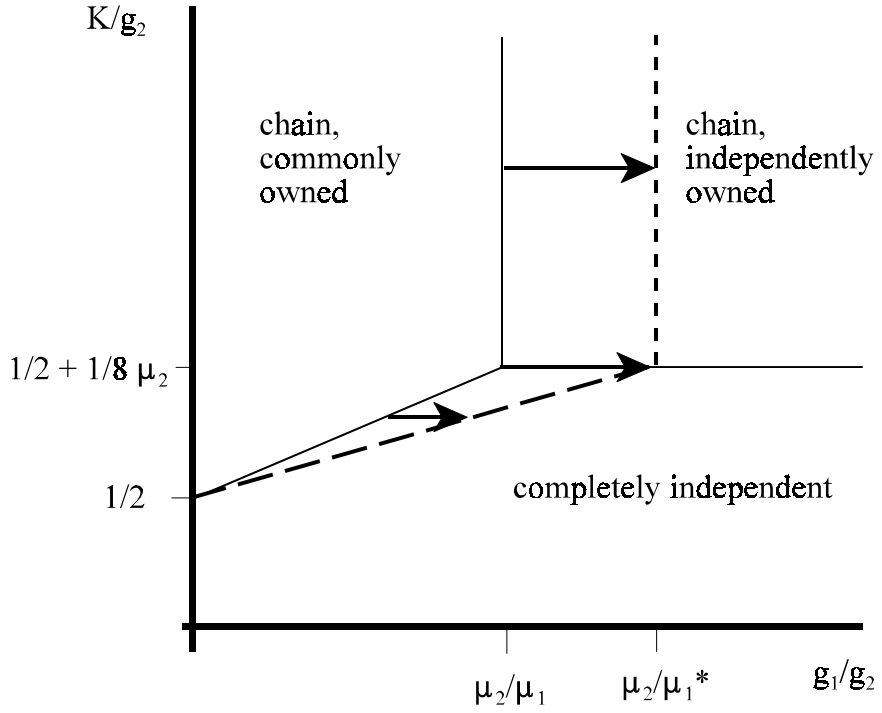


Figure 3
Increasing the Contractibility of e_2

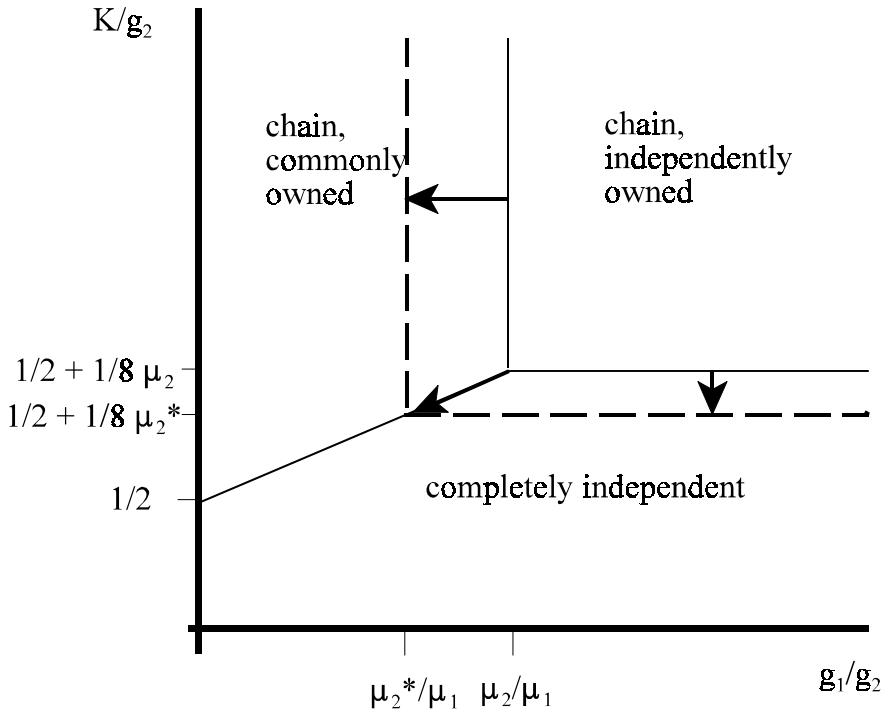
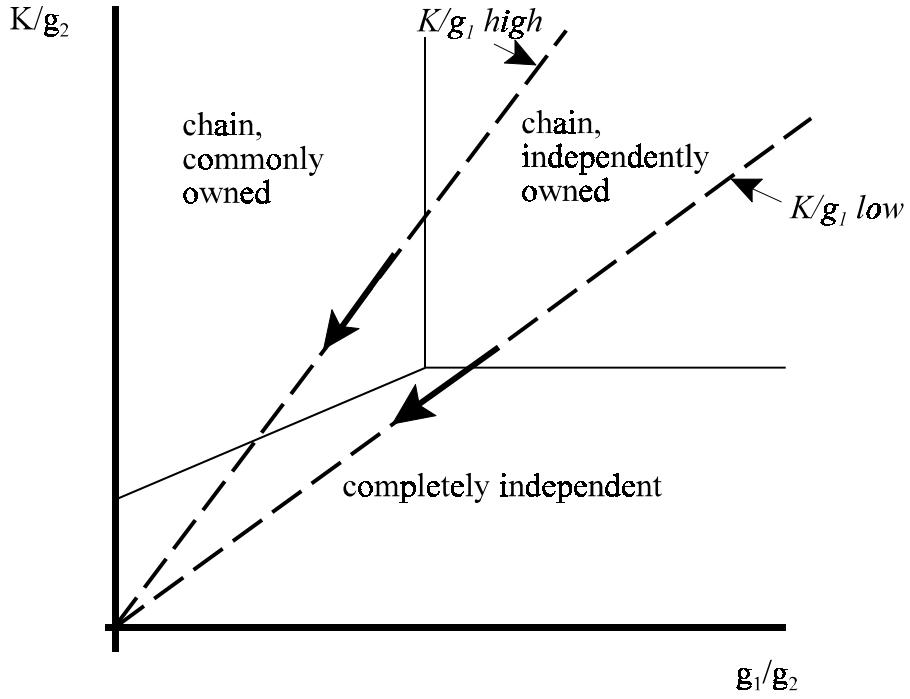


Figure 4
Increasing g_2



Dotted lines indicate expansion paths with respect to g_2. Arrows indicate the effect of increasing g_2.

References

Bain, Joe S., *Barriers to New Competition*, Cambridge, Harvard University Press, 1956.

Baker, George P., and Thomas N. Hubbard, "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," NBER Working Paper No. 7634, April 2000.

Baker, George P., and Thomas N. Hubbard, "Empirical Strategies in Contract Economics: Information and the Boundary of the Firm," *American Economic Review*, 91:2 (May 2001), pp. 189-194.

Brickley, James A., James S. Linck, and Clifford W. Smith, Jr., "Boundaries of the Firm: Evidence from the Banking Industry," University of Rochester, January 2000.

Bureau of the Census, *1992 Census of Service Industries: Establishment and Firm Size*, Washington, G.P.O., 1995(a).

Bureau of the Census, *1992 Census of Retail Industries: Establishment and Firm Size*, Washington, G.P.O., 1995(b).

Bureau of the Census, *1997 Economic Census. Real Estate and Rental and Leasing: Establishment and Firm Size*, Washington, G.P.O., 2000.

Chandler, Alfred E., *The Visible Hand: The Managerial Revolution in American Business*, Cambridge, Belknap, 1977.

Chain Store Age Executive, "Blockbuster POS Focuses on Three Goals," January 1990, p. 88.

Chakravaty, S., "Give 'Em Variety," *Forbes*, May 2, 1988, p. 54.

Dixit, Avinash K., and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67:3 (June 1977), p. 297-3xx.

Grossman, Sanford, and Oliver Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94:4 (August 1986), pp. 691-719.

Holmstrom, Bengt, "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 15:1 (April 1999), pp. 74-102.

Holmstrom, Bengt, and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7:0 (Special Issue 1991), pp. 24-52.

Holmstrom, Bengt, and Paul Milgrom, "The Firm as an Incentive System," *American Economic Review*, 84:4 (September 1994), pp. 972-91.

Lamoreaux, Naomi R., *Insider Lending: Banks, Personal Connections, and Economic Development in Industrial New England*, Cambridge University Press, Cambridge, 1994.

Love, John F., *McDonald's: Behind the Arches*, New York, Bantam, 1995.

Keefe, P., "Video Rental King Loads in Automation," *Computerworld*, February 15, 1988, p. 47.

Maness, Robert, "Incomplete Contracts and the Choice Between Vertical Integration and Franchising," *Journal of Law, Economics, and Organization*, 31(1996), p. 101-115.

Paikert, C., "Tuning In to the Home Video Explosion," *Chain Store Age Executive*, May 1982, p. 33.

Peterson, Mitchell A., and Raghuram G. Rajan, "Does Distance Still Matter? The Information Revolution in Small Business Lending," Kellogg Graduate School of Management, April 2000.

Salop, Steven C., "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 10: 1 (Spring 1979), p. 141-156.

Shepard, Andrea, "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing," *Rand Journal of Economics*, 24:1 (Spring 1993), pp. 58-77.

Spence, A. Michael, "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 66:2 (May 1976), p. 407-14.

Sutton, John, *Sunk Costs and Market Structure*, Cambridge, MIT Press, 1991.

Wiener, T., "The Video King and I," *Washington Post*, August 18, 1991.