

The economics of franchise contracts [☆]

Benjamin Klein

University of California, Los Angeles, Department of Economics, 405 Hilgard Avenue, Los Angeles, CA 90024-1477, USA

Revised 2 May 1995

Abstract

An incentive problem exists in franchise relationships because of the failure of franchisees to take account of franchisor profit. Franchise contracts ameliorate this malincentive not by specifying a proxy for desired franchisee performance, but by creating a premium stream that facilitates a self-enforcing agreement. The structure of credible commitments within this self-enforcing arrangement is elucidated, with initial franchisee investments shown to serve no performance guaranteeing purpose. Franchisors do not demand large initial lump sum payments from franchisees because doing so makes it more difficult to terminate franchisees for nonperformance. Franchisors use vertical integration when the premium necessary to assure franchisee performance is large.

JEL classification: L14; L22

Keywords: Franchise contracts; Self-enforcement commitment; Vertical integration

0. Introduction: What is franchising?

The three primary economic questions concerning franchising are: (1) why do transactors use the franchising form?; (2) what determines the particular contract terms chosen by transactors as part of their franchise contract?; and (3) what determines whether an outlet is owned and operated by the franchisor or owned and operated by an independent franchisee? I personally find questions two and three, the choice of contract terms and the

[☆] I am grateful to Kevin Murphy for many discussions of these issues over the past decade.

incidence of vertical integration, more interesting than question one why franchising. Franchising is just one of many ways a firm can choose to distribute its product. The essential economic rationale for franchising is that it permits transactors to achieve whatever benefits of large scale may be available in, for example, brand name development and organizational design, while harnessing the profit incentive and retailing effort of local owners. However, a similar economic rationale can be applied to distribution arrangements more generally and explains why manufacturers often leave the retailing of their products to independent, non-franchised retailers. For example, H.J. Heinz takes advantage of the significant economies of scale in the manufacture, distribution, and advertising of Heinz ketchup while leaving the retailing of their ketchup to independent grocers.

The legal definitions of franchising do not provide us with clear economic criteria to distinguish franchising from other forms of distribution. For example, the Federal Trade Commission Franchising Rule ¹ defines franchising by the presence of three factors: (i) distribution of goods and services associated with the franchisor's trademark; (ii) exercise of significant control over, or giving of significant assistance to, the franchisee by the franchisor; and iii) payment by the franchisee to the franchisor of at least \$500 before expiration of the first six months of operation of the franchised business. ² The first factor obviously is very general and would fit most distribution arrangements, while the third factor would eliminate franchises where the manufacturer does not charge a royalty, but instead earns a profit on product sales, such as automobile franchises. The key distinguishing economic element appears to be the second factor, the degree of control exercised by the franchisor over its franchisees. However, the degree of control exercised by a manufacturer over a retailer is often as great or greater than the control exercised by a franchisor over a franchisee.

For example, consider the sale of perfume in a department store. A perfume manufacturer may control the department store very closely, including hiring and training its own employees to work in the store to assist in the promotion of the product. Another example of close control is a clothing manufacturer, such as Ralph Lauren, that may build out separate department store space to its own specifications and at its own expense. Ralph Lauren then essentially rents this space by making a per unit time and/or percentage of sales payment to the store. These cases are clearly far along the continuum with regard to how much control is being exercised by a

¹ "Disclosure Requirements and Prohibitions Concerning Franchising and Business Opportunity Ventures", 16 C.F.R. §§436.1-3 (1988).

² 16 C.F.R. §436.2(a).

manufacturer over the retailing of its product, but we would not call them franchising.

The reason the department store is not considered a franchisee of the perfume or clothing manufacturer, in spite of how much control is being exercised, is primarily because there is a lack of exclusivity in the manufacturer-retailer relationship. Exclusivity is emphasized in the Department of Commerce definition of franchising as those arrangements that “concentrate on one company’s product line and to some extent identify their business with that company.”³ Exclusivity also fulfills the “community of interest in the business” criterion used in a number of State statutes to define the existence of a franchising relationship.⁴

The Department of Commerce makes a further distinction between franchising arrangements that are “business format franchises” as opposed to “product or tradename franchises”. The former covers “not only the product, service, and trademark, but the entire business format itself — a marketing strategy and plan, operating manuals and standards, quality control, and continuing two-way communication”.⁵ These are the highly standardized exclusive retailing arrangements associated with fast food franchises, such as McDonald’s, that are often the only examples of franchising considered by economists. However, it is not useful to limit the study of franchising to business format franchises. There is little economic difference between a Baskin-Robbins “business format” franchisee and an automobile dealer “trademark-product” franchisee. Both franchisees sell the franchisor’s product at a designated location under particular franchisor specified conditions. Business format franchising arrangements may, in general, entail a greater degree of franchisor control over the retailing operation. However, some automobile manufacturers, such as Lexus, now control the appearance of their showrooms as closely as Baskin-Robbins controls its outlets. Both arrangements essentially accomplish the same thing — retailers distribute products supplied by a manufacturer which also designs the national sales organization and promotional campaign.

³ U.S. Department of Commerce, *Franchising in the Economy* (1988), p. 1.

⁴ An example of such a statute is the Wisconsin Fair Dealership Law, Wis. Stat. Ann. §§135.01–.07 (West 1974 and Supp. 1982–83). The concept of “community of interest” is vague and extremely broad, with some courts holding that merely setting performance standards in a dealer agreement is sufficient to demonstrate the manufacturer’s “community of interest” with the dealer in the business and, hence, the existence of a franchise relationship. The recent 6th Circuit decision in *General Aviation, Inc. v. Cessna Aircraft Co.*, 915 F.2d 1038 (6th Cir. 1993) which held that the non-renewal of a distributor of Cessna aircraft was a violation of Michigan’s franchise law regarding unfair non-renewals is an example of this trend in the law. Also see *Boatland, Inc. v. Brunswick Corp.*, 558 F.2d 818, 824 (6th Cir. 1977) and *Van v. Mobil Oil Corp.*, 515 F. Supp. 487, 491 (E.D. Wisc. 1981).

⁵ U.S. Department of Commerce (1988), p. 3.

Whatever measures of control, exclusivity and standardization one decides to use as the basis by which to label a distribution arrangement as franchising, the above discussion suggests that one is likely to find a continuum of contract arrangements along each of these dimensions. Any sharply drawn lines are essentially arbitrary. Moreover, as we shall see, the fundamental economic forces that underlie franchising relationships, such as the incompatible incentives between the franchisor and franchisee with regard to promotion or marketing effort, are present in all distribution arrangements. Therefore, much of the economic analysis and many of the examples that follow are applicable to all distribution arrangements.

The paper consists of four sections. The first section describes why a franchisee's incentive to perform does not necessarily coincide with what would maximize the franchisor's profit. Incentive incompatibility between a franchisor (or manufacturer) and its franchisees (or distributors) is shown to be much more general than the problem of free riding on a common brand name that has been emphasized in the franchising literature. The second section outlines the role of contract terms in handling incentive incompatibility problems. Franchise contract terms are seen not as an alternative to self-enforcement, but as a complement that facilitates self-enforcement. The third section discusses the structure of credible commitments and, in particular, the role of initial franchisee investments in assuring performance of both the franchisee and the franchisor. The confusion of initial franchisee investments with "collateral performance bonds" that underlies much previous analysis is clarified. The fourth section uses the framework developed in the first three sections to outline the conditions under which a franchisor would find it more efficient to vertically integrate and operate its own outlets.

1. Franchisor and franchisee incentives do not coincide

The crucial economic fact that underlies franchising contracts is that the incentives of the transacting parties do not always coincide. It is for this reason that franchisors write contracts that attempt to control franchisee behavior. The economic literature on franchising concentrates on four types of franchisee behavior that must be controlled. One type of behavior that has been analyzed at great length is the free riding incentive created when franchisees jointly use a common brand name.⁶ In general, when franchisees use a common brand name, each franchisee can reduce its costs by reducing the quality of the product it supplies without bearing the full consequences of doing so. Because a reduction in quality has the effect of reducing the future

⁶ See, for example, Rubin (1978), p. 228 and Klein (1980), p. 358–59.

demand facing all franchisees using the common name, not just the future demand facing the individual franchisee who has reduced quality, the incentive for individual franchisees to supply the desired level of quality is reduced.⁷

Another commonly recognized case in which incentives do not coincide is when the franchisee provides some pre-purchase service that consumers can get free of charge at a full service franchisee before purchasing the product at a free-riding franchisee who does not provide service. This is referred to as the “special services” free-riding problem.⁸ For example, during the 1960s automobile manufacturers and full-service franchised automobile dealers experienced this problem with regard to discount “book dealers”, who sold cars out of catalogues without providing a showroom, inventory or sales staff.⁹

A third type of case discussed in the economic literature in which incentives do not coincide occurs when the franchisee possesses some power over price, perhaps due to the franchisor’s grant of an exclusive territory to the franchisee. This is the commonly recognized “successive monopoly” or “double marginalization” problem that is also sometimes claimed to exist in automobile franchising.¹⁰

However, the franchisee malincentive problem is much more general and pervasive than these three types of problems would suggest. In particular, a significant franchisee malincentive problem is likely to exist even when consumers do not rely on a brand name that is jointly used by multiple

⁷ While this externality or free riding problem is present in franchising arrangements because of the use of a common brand name, the problem is not unique to franchising. A similar problem may exist in the retailing of any product where retailers can influence the quality of the product they distribute and where consumers do not hold the individual retailer entirely responsible for the lower quality product they receive. In such a situation each retailer has an incentive to free ride on the manufacturer and group of retailers. For example, such a free-riding problem existed in the retailing of Coors beer, where each retailer had an incentive to save costs by not properly refrigerating the beer, resulting in a decrease in the future demand facing the manufacturer of Coors and all retailers of Coors. See Klein and Murphy (1988).

⁸ See Telser (1960). Since the Sylvania decision (*Continental T.V., Inc. v. GTE Sylvania, Inc.*, 975 S.Ct. 2549 (1977)) this is the distribution malincentive that has been emphasized by the courts, largely to the exclusion of all other distribution malincentives.

⁹ See *United States v. General Motors Corp.*, 384 U.S. 127 (1966).

¹⁰ See, for example, Smith (1982). Smith estimates a more than 9 percent successive monopoly price distortion in states where an exclusive territory has been legislatively expanded by “relevant market area” statutes (that prevent automobile manufacturers from freely adding new dealers into an area with established dealers). Such a large increase in automobile prices (amounting to about \$12,000 in terms of current average transaction prices), however, seems unlikely given the incentive consumers have to shop more than one dealer in spite of the grant of an exclusive territory. Eckard’s (1985) estimate of the effect of these statutes on price of less than one percent is likely to be closer to the true successive monopoly distortion.

franchisees, or when consumers do not free ride on services provided in full-service outlets before purchasing in low-service discount outlets, or when the franchisee does not possess any market power. All that is necessary for a malincentive problem to exist is that (1) the franchisee control the supply of some inputs that influence the demand for the franchisor's product, such as marketing effort; and that (2) the price at which the franchisor sells its product to franchisees is greater than its marginal cost. (I am assuming that franchising arrangements consist of a franchisor selling a product to franchisees who then resell the product to the public, as, for example, in the Baskin-Robbins type franchise system.)¹¹ Under these circumstances the franchisor has an incentive to sell additional units of its product and, therefore, desires that the franchisee supply additional quantities of marketing services, but the franchisee does not take account of the profit earned by the franchisor on incremental sales in determining how much marketing service to supply. This fourth type of malincentive is especially important when the franchisor's price-marginal cost gap is large and when the marketing services supplied by franchisees have a large influence on the franchisor's demand.

The example used in Klein and Murphy to illustrate this malincentive was the selling of a brand name perfume by a department store.¹² This example clearly illustrates the economic forces because the perfume manufacturer's margin is extremely high and because it is obvious that the perfume manufacturer wants the department store to supply a great deal of marketing services, such as prime shelf space and salespeople providing product demonstrations. However, similar economic forces are present in franchising arrangements involving the distribution of differentiated products, such as automobiles or soft drinks. The key economic fact is that franchisees, in determining how much marketing services to supply, do not take account of the effect of these marketing services on increased manufacturer sales and profits.

For illustrative purposes we can assume that the wholesale price of a product is set by the manufacturer or franchisor at \$4.50 and the competitive retail price is \$5.00, where the 50 cent differential between wholesale and retail price covers the competitive costs of retailing by franchisees. It is then

¹¹ This assumption makes franchising explicitly analogous to retailing and permits us to more clearly consider the franchisor's wholesale price-marginal cost gap as a key element of the economic analysis. In the McDonald's-type franchise arrangement, where an explicit product is not sold to the franchisee, the franchisor is also selling something, namely its brand name and the procedures which the franchisee must use to produce the final product and conduct its business. These more intangible franchisor products are likely to have an even lower marginal cost than in the franchising arrangements where a product is sold by the franchisor to franchisees and, therefore, the following economic analysis has greater applicability.

¹² Klein and Murphy (1988), pp. 282–285.

assumed that there is a potential consumer who values the product at only \$4.00, but who can be convinced to purchase the product if marketing services are supplied that have a cost of \$3.00. That is, an expenditure of \$3.00 will increase the consumer's valuation of the product by \$1.00. Obviously, it will not pay for the franchisee to supply these marketing services since the 50 cents the franchisee earns will only cover the normal costs of retailing, not the extraordinary marketing costs of \$3.00. However, if the franchisor's marginal cost of production is less than \$1.50, it would pay for the franchisor to see that these marketing services are supplied.

If the franchisor were vertically integrated into retailing, it would supply the marketing services. Alternatively, if the franchisor could compensate the franchisee for the cost of providing the marketing services, it would do so. But the franchisee, by itself, does not have an incentive to supply the services. Consumers will not demand and pay for the marketing services because the services are aimed primarily at marginal consumers who must be convinced to purchase the product. Therefore, the franchisee cannot increase the price of the product to reflect the cost of supplying the services, but must supply the services at a zero price. Supply of the marketing services can be thought of as a form of price discrimination since it only reduces the effective price to marginal consumers while the services are paid for by all consumers in the average price.

Sales effort at an automobile dealership would seem to fit this framework. Another example that fits the analysis is Monsanto's distribution of its agricultural herbicides.¹³ Monsanto wanted its distributors to hire and train additional sales staff to actively demonstrate the technical features of its new product and established minimum resale prices as the mechanism by which their distributors would be compensated for this additional marketing effort. An individual distributor, Spray-Rite, failed to hire and train the additional sales staff necessary to adequately promote Monsanto's product. However, the standard special services form of consumer free riding does not appear to have been practiced by Spray-Rite. Spray-Rite did not sell to individuals who first obtained the marketing services from another full-service distributor. Instead, Spray-Rite sold primarily at a discount to knowledgeable, large-volume customers who did not require the marketing services. While these customers did not obtain services from other distributors, they did take advantage of the fact that Monsanto assured distributors a sufficient margin on sales so that distributors would supply marketing services free of charge to customers that required such services. By Spray-Rite granting knowledgeable customers discounts and inducing them to switch their purchases from full-service distributors, they made it impossible for full-service distributors

¹³ *Monsanto Co. v. Spray-Rite Service Corp.*, 104 S.Ct. 1464 (1984).

to cover the costs of providing marketing services to the marginal customers. Rather than special services free riding, Spray-Rite and its customers were free riding on an implicit “price discrimination” marketing arrangement set up by Monsanto.¹⁴

In this context franchisee marketing services should be thought of very broadly to include any franchisee inputs that may influence the demand and hence the profit of the franchisor. For example, consider the question of the optimum number of franchise outlets in an area. A larger number of outlets may increase the demand for the franchisor’s product and, therefore, the franchisor’s profit. But an increase in the number of outlets will reduce the demand and profit of already established franchisees. Therefore, from an individual established franchisee’s point of view, additional outlets placed in its market area may represent unfair “encroachment” on its business.¹⁵

However, it is important to recognize that even if the externality between new and established franchisees were internalized, the franchisor’s interest in additional franchisees does not coincide with the established franchisees’ interests. For example, even if the established franchisee were granted the right to sell new outlets in its territory, so that there were no problem of a

¹⁴ Similarly, most examples of resale price maintenance cited in the economic and legal literature do not fit the special services free-riding analysis. Ippolito’s [1988] study of past Federal Trade Commission resale price maintenance litigation finds that a special services theory is consistent with the use of resale price maintenance in 65 percent of the cases. However, Ippolito did not examine the details of the marketing arrangements in these cases and only considered the types of products involved. In particular, if the product was complex, infrequently purchased, had a fashion element, or was sold by a new entrant, she concluded that the special services theory was a “potential explanation” for resale price maintenance. It is surprising that at least one of these conditions was present in only 65 percent of the cases. Moreover, when many of these cases where the special services theory is considered to be a potential explanation are examined more closely, they are found not to fit the theory. For example, all the branded clothing resale price maintenance cases (such as *FTC v. Levi Strauss & Co.*, 92 F.T.C. 171 (Consent Order entered July 12, 1978), *F.T.C. v. Jonathan Logan, Inc.*, 94 F.T.C. 159 (Consent Order entered July 25, 1979), or *F.T.C. v. Palm Beach Co.*, 98 F.T.C. 51 (Consent Order entered August 4, 1981)) fit the Ippolito criteria because the products have a fashion element. But the point-of-purchase sales promotion for such products are not special services that consumers are likely to obtain free of charge before purchasing the product at a discount store. Much of the marketing effort by retailers in these cases is devoted to convincing marginal consumers, who would not have purchased the product absent the marketing effort, to purchase the product at that particular point in time. Manufacturers of these types of products often face negatively sloped demands and high price-marginal cost differentials and, therefore, find it in their interest to move down their demand curves by having retailers actively promote their products.

¹⁵ This is the motivation for laws in many states that restrict the placement of new automobile dealerships near established dealers that are selling the same make of automobile. In addition, much litigation has occurred against non-automobile franchisor “saturation” of established franchisees’ “relevant market areas”. Perhaps the most notorious case on this question is *Photovest Corp. v. Fotomat Corp.*, 606 F.2d 704 (7th Cir. 1979), which is discussed below (fn. 25).

franchisor taking advantage of the franchisee's investment in building up the territory or of not considering the lost sales of the established franchisee, the established franchisee would not sell the correct number of new outlets from the franchisor's point of view. The franchisor and established franchisee do not have the same incentive to increase outlets because the established franchisee does not take account of the extra profit earned by the franchisor on any additional sales produced by the extra outlets. While the franchisor gains from additional outlets, the extra sales at the retail level may not cover the cost of the extra outlets. It is this problem of inadequate outlets (or of inadequate marketing effort more generally) that must be handled in the franchise contract.

2. The role of contract terms

Franchise contracts are designed in part to control the malincentive problems that exist between franchisees and franchisors. In examining how particular franchise contract terms assure desired franchisee performance in the face of these problems, it is useful to distinguish between two different ways in which contract terms operate. First of all, contract terms may create the correct incentive for the franchisee to perform by specifying desired performance or some proxy for desired performance to be enforced by the court. For example, the contract may specify a desired level of franchisee marketing effort or, more realistically, specify some proxy for the desired level of marketing effort, such as the number of salespeople that must be hired. Contractually specified proxies for desired franchisee performance also may be more indirect. For example, the contract may specify an exclusive territory for each franchisee, knowing that this will increase the probability of consumer repeat sale and thereby create an increased incentive for franchisees to perform as desired. In all these cases of contractual specification, whether performance is specified directly or indirectly, it is the threat of court enforcement of the contract term itself that is assumed to be sufficient to produce desired franchisee behavior.

This first way in which franchise contracts may operate is generally not sufficient to assure franchisee and franchisor performance because it is not possible to specify in a legally enforceable document all elements of desired behavior nor to be able to use perfect proxies for desired behavior. Specifications of or proxies for desired behavior are necessarily imperfect because desired behavior is extremely complex and difficult to measure, with contract specifications only incompletely covering all the many elements of performance and all the potential unforeseen future contingencies. This does not mean that performance is unknowable. The franchisor, for example, may know fully after the fact whether the franchisee has performed satisfactorily or not. It merely may be impractical to put in a written contract the way in

which the franchisor learns about complex franchisee behavior, such as the effort and energy the franchisee has devoted to marketing. As a result, attempts to define desired performance in a written contract will lead transactors not to supply the desired behavior but to maximize against the imperfect measures of performance in the contract.

Therefore, rather than relying solely on imperfect contracts to specify desired behavior or some proxy for desired behavior to be enforced by the court, a franchise contract may be designed to assure performance in a very different way, namely by facilitating a self-enforcement mechanism. Court-enforced contract terms operate in this context by creating sufficient franchisee rents so that the threat of termination of the relationship by the franchisor gives the franchisee sufficient incentive to supply the desired behavior. This mechanism requires the franchisor to actively monitor franchisee behavior, in the sense of determining if the desired behavior is supplied and then terminating the franchisee if the desired behavior is not supplied. It is the threat of termination, rather than the use of the court to enforce the written terms of the contract alone, that produces the incentive on the part of the franchisee to perform.

To understand how this self-enforcement mechanism operates, let us denote the present discounted value of the extra profit that a franchisee could earn by non-performance over the short-term (before being detected and terminated by the franchisor) as W_1 . This represents how much a franchisee can reduce its costs in the short-term by not performing as the franchisor desires. W_1 is determined in part by the resources the franchisor devotes to monitoring franchisees and detecting non-performance. The franchisor will spend increased dollars on franchisee monitoring until an extra dollar spent equals the decrease in W_1 that is brought about by the expenditure. This is because, as we shall see, a decrease in W_1 decreases the required premium stream that the franchisor must pay the franchisee to assure performance.

If a self-enforcement mechanism is to operate, contract terms must create sufficient rents or a future “premium stream” which, combined with the threat of termination, produce the incentive to perform. Let us denote the present discounted value of this premium stream that a franchisee could earn over the long-term by performance as W_2 . The self-enforcement contract will assure franchisee performance if the contract creates a future expected premium stream for the franchisee if it performs as desired that is equal to or greater than the short-run gains to the franchisee if it does not perform as desired.

$$W_2 \geq W_1. \quad (1)$$

The first purpose of contract terms — to define desired behavior or some proxy for desired behavior to be enforced by the court — can be thought of

analytically as using contract terms to decrease W_1 . The second purpose of contract terms — to create sufficient rents which when combined with the threat of termination produce the incentive to perform — can be thought of analytically as using contract terms to increase W_2 . Since a contract cannot perfectly define desired behavior, i.e., decrease W_1 to zero, a role must remain for self-enforcement, or W_2 . The franchisor can be thought of as setting court-enforced contract terms to minimize W_1 and then structuring the contractual relationship so that W_2 is greater than W_1 .¹⁶

What may be somewhat confusing is that the same franchise contract term can simultaneously serve both purposes. For example, an exclusive territory can decrease W_1 by increasing the probability of consumer repeat sale and, therefore, internalize franchisee externalities associated with decreasing quality. But, more importantly, an exclusive territory also creates a valuable asset and, hence, a future premium stream, W_2 , that the franchisee can lose for non-performance. Similarly, resale price maintenance can decrease W_1 , by decreasing the ability of the franchisee to lower price and expand sales in the short-run if it decides not to perform; but, once again more importantly, resale price maintenance also can create a franchisee premium stream and, therefore, increase W_2 .¹⁷

This view of franchise arrangements, where franchisee performance is assured by franchisor monitoring and by the threat of loss of a future expected premium stream, has been presented in a number of articles.¹⁸ However, the role of contract terms in facilitating this self-enforcement mechanism is not fully appreciated. Unfortunately, most economists still focus solely on the first purpose of contracts, namely the role of contract terms in creating the correct marginal incentives on a contractually specified measure of (or proxy for) performance, and ignore the second purpose of contract terms, the creation of rents sufficient to make the relationship self-enforcing.

¹⁶ More precisely, contract terms should be thought of as minimizing the probability that W_1 is greater than W_2 , with the actual values of W_1 and W_2 depending upon the particular ex post market conditions that develop. For example, the growth of demand in an individual franchisee's area may decline unexpectedly, reducing W_2 below W_1 . Contract terms, therefore, define the "self-enforcing range" of the franchise relationship, or the range of likely ex post market conditions where the gain to the franchisee from performing will be greater than the gain for not performing. And the franchisor may decide to intentionally set the expected level of W_2 above the expected level of W_1 to take account of the likelihood of these stochastic shocks. However, there will always remain some ex post market conditions where W_1 would be greater than W_2 . See Klein (1995).

¹⁷ See Klein and Murphy (1988).

¹⁸ See, for example, Rubin (1978), Klein (1980), Mathewson and Winter (1985) and Klein and Murphy (1988).

For example, this is the primary deficiency in the Telser (1960) special services analysis of resale price maintenance. A high fixed minimum price, by itself, will not create the correct incentive on franchisees to supply the desired services because, contrary to Telser's assumption, there is more than one dimension of franchisee non-price competition. Therefore, franchisees will continue to have an incentive to free ride on full-service franchisees by supplying other (non free-rideable) services to consumers that first obtain the desired special services from a full-service franchisee. In addition, franchisees will not have an incentive to supply those elements of desired franchisee service that cannot be detected by consumers pre-purchase (e.g., the quality of meat used to make a hamburger). Hence, to assure franchisee performance franchisors have to do more than merely set the retail price; they must create an arrangement where franchisee profit is not competed away and they must also actively monitor franchisee performance.¹⁹

Detailed empirical analyses of how actual franchise contracts serve both contract purposes, the creation of direct incentives and the creation of economic rents, are necessary in order to expand our knowledge of franchising arrangements. Studies must be undertaken to document why a franchisee premium stream is required, i.e., what elements of franchisee performance are difficult to specify and unlikely to be supplied in optimal amounts, and how the franchise contract creates the premium stream. For example, the contract may use exclusive territories, or more generally the number and spacing of outlets, or use minimum resale price maintenance (by the use of, say, product allocations) as alternative ways to create the franchisee premium stream. What will be used to create a franchisee premium stream will depend upon the facts of the particular arrangement, including the type of product franchised. For example, a franchisor of a low price convenience item that has spent a considerable amount of money on national advertising to create a demand for its particular product, such as McDonalds, may not find it

¹⁹ See Klein and Murphy (1988) for a fuller discussion of these issues. Telser's criticism of the Klein and Murphy analysis (Telser (1990)) is based upon a claim that it is not possible to have a Nash equilibrium where the franchisor monitors franchisees and where franchisees perform as desired because if franchisees never cheat, then the franchisor will not monitor; but then franchisees will cheat. However, this is just an artificial result of Telser's specific model. If franchisees can observe the degree of franchisor monitoring prior to deciding whether to cheat (for example, one decides whether to rob a bank only after observing policing efforts and estimating the probability of detection), then an equilibrium with no franchisee cheating can be sustained. Telser also criticizes the Klein and Murphy analysis by claiming that the self enforcement mechanism is inefficient compared to the franchisor merely collecting damages from franchisees that do not perform. However, the explicit assumption underlying Klein and Murphy (and also implicitly underlying Telser's original analysis) is that the franchisor cannot specify desired franchisee behavior in a court-enforceable contract and, therefore, the franchisor cannot collect damages in court.

necessary to grant a large market area in order to generate a franchisee premium stream. The problem in this type of case actually may be one of holding the price down, monitoring franchisee pricing and enforcing *de facto* maximum resale price maintenance with, for instance, suggested retail prices to avoid the “double marginalization” problem.²⁰

On the other hand, for a product such as automobiles, where consumers generally can delay purchase of the product and where it pays consumers to search over a much wider area than they search for a hamburger, the exclusive territory that would be required to produce a particular percentage of sales premium may be greater than would be optimal from a consumer convenience or manufacturer marketing standpoint. This would certainly be the case if franchisees could purchase as many automobiles as they wished from the manufacturer and then advertise a low price over a broad market area. The resulting equilibrium number of franchisees would be much smaller than would be profit-maximizing from the manufacturer’s point of view. In this case *de facto* minimum resale price maintenance in one form or another (e.g., product allocations) may be used by the manufacturer.

Perhaps a clearer example of the use of minimum resale price maintenance to create extra outlets as a profit-maximizing marketing arrangement is the franchising arrangements adopted by candy manufacturers, such as See’s Candy.²¹ Obviously no special services-type free riding is involved here. But without fixed minimum retail prices, one See’s franchisee could lower price and get the price-sensitive customers who know they want to purchase the candy and are willing to search for it, leaving the other franchisees to handle primarily the impulse consumers who decide to purchase the candy only when they see the outlet. As with the Monsanto arrangement, the franchisor in the See’s case adopted a marketing arrangement where all customers

²⁰ The franchisor generally can control any successive monopoly problem. For example, if the franchisor ships a product to the franchisee and demand is fairly predictable, the franchisor can use minimum sales targets which the franchisee must meet to retain the franchise or to receive some particular benefit. Alternatively, the franchisor can fix a maximum price by the use of advertised “suggested” retail prices at “participating” franchisees and impose sanctions on franchisees that do not participate. A particularly creative way for a franchisor to set maximum prices is illustrated in *Eastern Scientific Co. v. Wild Heerberg Instruments, Inc.*, 572 F.2d 883 (1st Cir.), cert. denied, 439 U.W. 833 (1978). The manufacturer in that case granted exclusive territories but permitted interterritory sales if they occurred at list prices. By creating a threat of sales from dealers in other territories, the manufacturer thereby set the list price as the effective maximum price that dealers could charge customers in their own territories. This arrangement was upheld as legal because it was considered less restrictive than a legal, pure exclusive territory arrangement.

²¹ The Federal Trade Commission brought resale price maintenance cases against Barton Candy (*F.T.C. v. Barton’s Candy Corp.*, 79 F.T.C. 101 (Consent Order entered July 21, 1971) and Russell Stover Candies (*F.T.C. v. Russell Stover Candies, Inc.*, 100 F.T.C. 1 (Consent Order entered July 1, 1982)).

subsidized the extra outlets. The extra outlets should be considered a form of price discrimination marketing that is necessary to get impulse consumers to purchase the product.

3. The structure of credible commitments

In order for the self-enforcement mechanism to work, the franchisor must be able to guarantee franchisees that they will receive a premium stream in the future if they perform as desired. However, franchisors can credibly commit to pay a future premium stream to franchisees only if paying the premium is cheaper for the franchisor than not paying the premium. If the franchisor does not pay the premium stream, it saves the cost of the premium payment. However, the franchisor then bears the additional cost of handling distribution with some non-franchising system, such as employees.²² The franchisor's cost disadvantage of operating the system completely with employees compared to franchisees, therefore, determines the ability of the franchisor to commit, i.e., the maximum level of the future premium stream that the franchisor can credibly promise to pay franchisees.

We can represent the franchisor's cost disadvantage of operating the system completely with employees compared to franchisees by $(C_e - C_f)$. As long as the present discounted value of this cost disadvantage to the franchisor of operating the system with employees is greater than the present discounted value of the promised premium stream to the franchisees, it will be in the franchisor's interest to pay the promised premium stream. That is because the franchisor's decision not to pay the premium also must lead the franchisor to adopt the more costly employee-run operation to avoid the franchisee non-performance that results from failure to pay the premium. Therefore, the franchisor's promise to pay a future premium of W_2 to franchisees will be a credible commitment if it fulfills condition (2).

$$W_2 \leq \sum_t \left[\frac{(C_e - C_f)_t}{(1 + r)^t} \right]. \quad (2)$$

At every point in time the present discounted value of the cost disadvantage of using a non-franchising organization must be greater than the present

²² We can assume that if a franchise system has been established, it is because such a system (which may include a certain fraction of employee-operated outlets) is the lowest cost method of distribution. Some evidence that labor costs are higher in employee-operated outlets than in franchised outlets is contained in Krueger (1991) and Shelton (1967). In addition to higher wage rates, employee managers, even those with profit sharing/bonus contracts, generally have a reduced incentive to perform. In part, this is because such managers are not residual claimants with rights to sell the franchise.

discounted value of the promised premium stream. If this condition is met, franchisees believe that they will receive the premium stream in the future because it is too costly for the franchisor not to pay the premium. If, on the other hand, the premium stream necessary to assure franchisee performance, W_2 , is greater than the discounted value of the cost advantage from franchising, franchising will not occur.

There is substantial evidence that franchisors have credibly guaranteed franchisees a premium stream sufficient to assure performance in many cases. For example, Mathewson and Winter (1985) interpret the existence of substantial queues for the right to become a franchisee in many franchise systems as evidence that potential franchisees expect to earn significant rents in these systems. Additional indirect evidence has been provided by Lafontaine (1992), who finds no significant negative correlation between royalty rates and franchise fees across franchise systems. The most convincing and direct evidence has been provided recently by Kaufman and Lafontaine (1994) in an excellent paper that carefully documents the financial return of McDonald's franchisees. Kaufman and Lafontaine conservatively estimate the capital value of the premium stream received by McDonald's franchisees at about half a million dollars per outlet on average.²³

An obvious question that Kaufman and Lafontaine attempt to answer is why franchisors generally do not demand initial lump sum fees from their franchisees equal to the present discounted value of the premium the franchisees expect to receive.²⁴ That franchisors could collect such large lump sums is irrefutably implied by the underlying economic logic. As we have seen, the franchisor sets up a system where the franchisee performs as the franchisor desires because the franchisee expects to receive a premium stream in the future, the present discounted value of which is greater than the extra short-run profit the franchisee would receive by not performing. If one claims that lump sums are not collected by franchisors because franchisees would not be willing to pay such lump sum fees, it is equivalent to claiming that franchisees do not expect to receive the premium stream and, therefore, that franchisees will not perform as desired. If the franchise system is working with franchisees performing as desired, then franchisees must expect to receive the premium stream and, hence, would be willing to pay an initial lump sum equal to its present discounted value. Failure of franchisors to demand sufficiently large lump sums implies that franchisors are giving something valuable away, a difficult thing for economists to

²³ Kaufman and Lafontaine estimate that the ex ante capital value, after payment of any initial fees or investments, ranges from \$300,000 to \$455,000 in 1982 dollars, which converts to a range of \$460,000 to \$700,000 in 1994 dollars. The ex post capital value, i.e., the discounted value of the premium stream excluding any initial fees or investments is, of course, larger.

²⁴ McDonald's franchisees currently pay an initial franchise fee of only \$45,000.

comprehend. Before solving this puzzle, we first examine the economic role of initial lump sum franchise fees.

3.1. The role of initial franchisee investments

First of all, it should be emphasized that the absence of an initial lump sum franchise fee does not imply a failure of the self-enforcement mechanism. Although the ability of the franchisor to make a credible commitment to pay a future premium stream, condition (2), also permits the franchisor to collect an initial lump sum equal to the capital value of the promised premium stream, whether the franchisee actually pays this lump sum or not has no effect on franchisee behavior. What the franchisee loses if it does not perform and is terminated is the discounted value of the expected future premium stream. Therefore, it is the expected future premium stream compared to the short-run gain from not performing, condition (1), that determines franchisee performance. Whether the franchisee paid an initial lump sum in the past is a sunk cost that is irrelevant to the franchisee's calculation of whether or not to perform now.

Franchisee payment of an initial lump sum also does not change the likelihood that the franchisor will perform as desired. Although one may think of the franchisor as more likely to cheat franchisees that have made initial investments by terminating them without cause and “grabbing” their investments, such franchisor behavior is independent of initial franchisee investments. The necessary and sufficient condition for such franchisor non-performance is solely the absence of condition (2). If the future premium stream that has to be paid by the franchisor to franchisees is greater than the present discounted value of the cost savings from operating the franchise system, then the franchisor will terminate the franchisees. Whether the franchisee has made an initial investment or not, what the franchisor gains by not performing and terminating franchisees unfairly is not having to pay the future premium stream. If, for example, franchisees have paid a lump sum equal to the present discounted value of the premium stream, franchisees can be thought of as having already paid the franchisor for the future premium stream. But once this lump sum is paid up front, the franchisor has the same gain from terminating franchisees as if the lump sum were not paid, namely not paying the future premium stream. And whether this gain outweighs the cost to the franchisor, namely the added costs of distribution, is also independent of whether franchisees have made an initial investment.

To sum up, only conditions (1) and (2) matter in determining franchisee and franchisor behavior. The franchisor will pay the premium to franchisees as long as condition (2) holds, namely the present discounted value of the cost savings from a franchising arrangement outweighs the present discounted value of the premium stream that must be made. And the franchisee

will perform as long as condition (1) holds, namely the present discounted value of the premium stream is greater than the short-run gain from not performing. Whether an initial lump sum payment is made by the franchisee does not alter either of these calculations.²⁵

A significant amount of confusion exists with regard to this analysis. Many discussions of the self-enforcement mechanism refer to franchisor requirements of initial franchisee investments, including lump sum payments, as equivalent to “collateral performance bonds”. These initial franchisee payments guarantee performance, it is asserted, because they serve as a “hostage” that will be lost upon termination. Unfortunately, I am also guilty of this error, having incorrectly claimed that:

“the franchisor may require an initial lump sum payment from the franchisee equal to this estimated short-run gain from cheating. This is equivalent to a collateral bond forfeitable at the will of the franchisor. The franchisee will earn a normal rate of return on that bond if he does not cheat, but it will be forfeited if he does cheat and is terminated”.²⁶

It is misleading to describe initial lump sum payments as collateral bonds that assure franchisee performance since it is always the future premium stream that determines whether a franchisee will perform or not perform. An

²⁵ As noted above (fn. 16), it makes more economic sense to think of conditions (1) and (2) in probabilistic terms, where transactors voluntarily entering into these relationships expect conditions (1) and (2) to hold but recognize that changes in market conditions may occur to force the relationship outside the “self-enforcing range” where the conditions will hold (Klein (1995)). For example, it may not be in the franchisor’s interest to perform and pay the premium stream to franchisees and, instead, to terminate franchisees when there is an unexpected increase in the discount rate or an unexpected decrease in the growth of demand (both of which imply a higher required premium to keep W_2 above W_1), or an unexpected decrease in the cost of vertical integration (which implies that the franchisor cannot credibly commit to pay the required premium stream). The Fotomat case (*Photovest Corp. v. Fotomat Corp.*, 606 F.2d 704 (7th Cir. 1979)), where Fotomat drove its franchisees out of business by oversaturating the market with additional company-operated outlets and by raising franchisee costs, appears to have been due to this latter effect, i.e., Fotomat’s recognition that it had lower costs of using employee-operated outlets. Fotomat, therefore, did not have the incentive or the ability to pay the required franchisee premium stream. However, because potential franchisees generally enter into their arrangements with a reasonable amount of knowledge and these changes in market conditions are infrequent, opportunistic franchisor problems are generally relatively rare. Further, when opportunistic franchisor problems do occur, we should recognize that, contrary to the decision in Fotomat, it is a contract problem not a monopolization problem.

²⁶ Klein (1980), p. 359. Similar mistaken statements have been made by, for example, Dnes (1993) and Williamson (1983) among others.

initial lump sum may usefully be considered a bond in the sense that franchisee payment of the lump sum is equivalent to paying the franchisor for the future premium stream up front, with the premium then merely representing an interest return on this initial payment. But a lump sum is certainly not collateral in the sense that it in any way assists in the operation of the self-enforcement mechanism.

Part of the confusion may be based on the way in which an equilibrium is sometimes reached, by initial franchisee investments in specific (nonfully salvageable) production assets that are lost upon termination.²⁷ However, even in this case we must be clear that it is the future return earned on these specific productive assets that assures franchisee performance, not the fact that the franchisees have made the specific investments. For example, if the franchisor had made the specific investments in the outlet, but the franchisee was earning the return from those investments, the incentive on the franchisee to perform would be the same, i.e., the fear of loss of the future return from these specific assets upon termination. Whether the franchisee makes the investment is irrelevant. While it may be easy to think of the loss to the franchisee of these specific assets upon termination, it is always the loss of the present discounted value of the future premium (which may include a future expected return on specific assets) compared to the short-run extra return from not performing that determines whether the franchisee performs.²⁸

Another element in the confusion that leads economists to think of initial investments, including lump sums, as “collateral bonds” is the fact that these investments appear to be returned to franchisees when they sell their franchise. Franchisees must be permitted to sell their franchise in order for the self-enforcement mechanism to operate. If the franchisee cannot sell the franchise for the present discounted value of the future premium stream in

²⁷ Since the normal return on these productive specific assets can be used to assure franchisee performance, W_2 should be defined to include this return and the premium then need not be above a normal rate of return by the full amount of the short-run gain from not performing.

²⁸ The usual example of forfeitable collateral put up by an individual when obtaining a bank loan, say a house used as collateral for a mortgage loan, is fully consistent with this analysis. The potential loss of the house, or the present discounted value of the future housing services from the house, is what motivates the individual to pay the interest on his mortgage loan. The only difference is that in the mortgage loan case, as opposed to the franchising case, the bank must demand the initial lump sum. The bank would not find it profitable to make the investment, i.e., to buy the house for the individual and threaten to take it away if the individual does not pay its mortgage. Although such a threat would be sufficient to assure the mortgage payments, the bank has no additional gains from trade, as does the franchisor, that would ever lead it to structure the relationship in this way.

its last period, it will not be in the franchisee's interest to perform in its last period.²⁹ However, when a franchisee sells its franchise, it does not just "get its money back". What the franchisee actually is selling is the discounted value of the future premium stream expected at that point in time; that is all a buyer is willing to pay and all a franchisee has to sell. This amount may be less than or greater than whatever initial lump sum payment may have been made by the franchisee. Even if the initial lump sum fee were small or non-existent, the franchisee can sell the franchise for the discounted value of the future premium stream, i.e., the full market price.

Initial investments also are irrelevant when the entire franchise system's time horizon is finite, i.e., when there is a last period where the franchisee cannot sell its franchise to anyone. To assure performance in this case, the franchisee must expect to receive a final "severance" payment from the franchisor in the last period. But, once again, it is the promise of severance, independent of whether a franchisee lump sum has been paid or not, that guarantees franchise performance. And the franchisor can credibly commit to pay this severance only if the cost to the franchisor in the future of not paying the severance is greater than the franchisor's gain in terms of saving the severance expense. The franchisor's incentive to pay the severance depends on the present discounted value of the cost that will be imposed on the franchisor in conducting other activities in the future (such as the creation and operation of other new franchise systems) compared to what the franchisor can save by not making the payment. The franchisor's incentive to pay the severance payment, therefore, will not depend upon whether the franchisee has paid an initial investment up front.

Because the capital cost to the franchisor of not making the severance payment (i.e., the franchisor's reputation) is limited, "ugly princesses" may be an important element of the solution in this finite time horizon type of case.³⁰ In particular, if the franchisor possesses an asset which it does not value but which the franchisee values sufficiently highly to induce franchisee performance, the franchisor promise of this asset as a severance payment to the franchisee in the last period is more credible. (Although the franchisor does not value the asset, the franchisor still can threaten not to give it to the franchisee in the last period; that is, it becomes a bilateral monopoly problem. But use of such an asset reduces the franchisor's required reputation capital or the capital cost imposed on the franchisor when not making the promised severance payment.) However, once again, it is not necessary

²⁹ This is "the last period problem". Because of rational expectations on the part of the franchisor and franchisee, the equilibrium will unravel and the franchisee will not perform in the first period. Therefore, to fulfill condition (1) and prevent franchisee non-performance the franchisor must permit the franchisee to sell its outlet in the last period.

³⁰ See the analysis of hostages in Williamson (1983).

that the franchisee make the investment in this asset and give it to the franchisor to be returned to the franchisee in the last period. The gains from trade may be sufficient to permit the franchisor to make such an investment, say in purchasing a painting from the franchisee that only the franchisee places any value on, and the system will operate.

3.2. Legal constraints on the self-enforcement mechanism

The analysis to this point of franchisor non-performance has assumed that there are no legal constraints on the franchisor's behavior. If, alternatively, there are legal constraints that place limits on the franchisor's ability to unfairly terminate franchisees, then initial lump sum payments made by franchisees may serve some economic purpose. In particular, the fact that franchisees have paid initial lump sums to the franchisor may make it more difficult legally for the franchisor to terminate franchisees without paying severance. Initial franchisee lump sums then may be a way to economize on franchisor reputation in solving the last period problem (and the problem of opportunistic termination of franchisees more generally). While initial franchisee investments are irrelevant in determining franchisee behavior or franchisor behavior in all the cases discussed above, initial franchisee investments may have a role in influencing the courts to prevent franchisor cheating. Therefore, in addition to the franchisor's obvious desire to collect an up-front payment from the franchisee before paying the franchisee the future promised premium stream, the franchisor also has an allocative reason to use a large initial lump sum as part of the franchise contract. The absence of large initial lump sum payments in franchise contracts then appears to be even more of a mystery.

Kaufman and Lafontaine (1994) answer the question of why franchisors do not demand an initial lump sum payment equal to W_2 , the maximum amount that their franchisees would be willing to pay, based upon franchisee wealth constraints. They maintain that the franchisees preferred by McDonald's do not have the assets or borrowing ability to raise the money that would be required for such large initial lump sum payments.³¹ This explanation is incredible. Surely potential franchisees can get a loan based upon the expected return, at least for a major part of the present discounted value of the expected future premium, when investing in a "blue chip" franchise such as McDonald's. If necessary, the franchisor could assist franchisees in obtaining this financing. Convincing evidence that financing is available is the fact that individuals who purchase outlets from retiring franchisees do not have a

³¹ Mathewson and Winter (1985) also rely on this assumption.

problem financing these purchases. Moreover, initial lump sums often are absent in cases where the prospective franchisees are relatively sophisticated and wealthy business people.³²

The answer to the puzzle of why large initial lump sum franchise fees are absent in franchise contracts may be related to the fact, noted above, that lump sums may place legal constraints on a franchisor's ability to terminate franchisees opportunistically. The opposite side of the coin is that lump sums also may place legal constraints on the ability of franchisors to terminate franchisees for non-performance. In particular, the legal framework under which franchise contracts are enforced makes it difficult for franchisors to use the termination-at-will sanction that underlies the self-enforcement mechanism.³³ While courts in a majority of states do not require the franchisor to show "good cause" or its equivalent in order to terminate or not renew a franchisee, contract law and the good faith obligation of the Uniform Commercial Code imposes a duty on a franchisor in all states for good faith and fair dealing.³⁴

Within this legal environment a number of courts have permitted franchisors to use the termination-at-will sanction. These courts have failed to find franchisee termination-at-will clauses unconscionable in part because the termination did not result in the franchisor either usurping funds to which

³² For example, in the Coors' marketing arrangement examined in Klein and Murphy (1988) Coors granted wholesale distributors relatively large, extremely valuable exclusive territories without demanding any initial franchisee fee at all, in spite of the fact that many of these distributors were already successful businesspeople. Distributors had to make significant investments in refrigerated warehouses and refrigerated trucks. However, these were not primarily Coors-specific investments; they were salvageable assets that would not have been lost upon termination. What would have been lost upon termination was the "extra" return earned on these assets, a return that was not from the assets, but from the grant (or gift) of the exclusive territory franchise.

³³ Specific statutes at the Federal level include the Automobile Dealer Franchise Act (15 U.S.C. §§1221–1225 (1982)) which permits automobile dealers to sue in federal court for damages caused by an automobile manufacturer's failure to act in good faith in terminating or not renewing them and the Petroleum Marketing Practices Act (15 U.S.C. §§2801–2806 (1982)) which prohibits oil company terminations or non-renewals of petroleum franchisees other than for certain specific reasons. At the State level a number of State franchise laws impose a "good cause" requirement for terminations or non-renewals, with some statutes also requiring the franchisor to give the franchisee notice of default and an opportunity to cure the defect. In addition, some statutes give terminated franchisees the right to injunctive relief. The effect of these provisions is to increase the franchisee's ability to not perform without being terminated, increasing W_1 and, thereby, the required premium. This is consistent with the evidence presented in Brickley, Dark and Weisbach (1991) which indicates that the extent of franchising is lower in states that require "good cause" for franchisee terminations.

³⁴ See Restatement (Second) of Contracts §205 (1979) and UCC Section 1–203 (1989).

the franchisee was entitled, or depriving the franchisee of income it had earned, or leading to a loss of the franchisee's financial investment.³⁵ If initial lump sum investments are made by the franchisee, these laws would likely require a refund of the lump sum upon termination. In fact, some of the statutes designed to protect franchisees against termination without "good cause" explicitly require a franchisor to compensate a terminated (or non-renewed) franchisee for certain defined categories of assets, and an initial franchisee-paid lump sum would be an obvious candidate for this type of refundable asset.³⁶ What these statutes fail to recognize is that termination, if it is to work in assuring contractually unspecified franchisee performance, must by necessity be "unfair" in the sense that the expected cost to the franchisee must be greater than the franchisee's expected short-run gain from not performing.³⁷ Therefore, legal attempts to control opportunistic terminations of franchisees entail the associated cost of making the self-enforcement mechanism more difficult to use.

Given the state of the law, the effectiveness of the self-enforcement mechanism may depend upon the absence of initial lump sums. If the franchisee paid an initial lump sum equal to the discounted value of the future premium stream up front which the law then required to be returned to the franchisee upon termination, the self-enforcement mechanism could not work. No matter how high the promised premium stream, a franchisee would always be better off not performing. In addition to the short-run gain from not performing, W_1 , the terminated franchisee would also receive at the time of termination the initial lump sum, which is equal to the discounted value of the premium stream, W_2 . Therefore, the return from not performing, $W_1 + W_2$, would always be greater than the return from performing, W_2 . A franchisee would never perform because it would not lose anything when terminated. On the other hand, if the franchisee does not pay the franchisor an initial lump sum fee, termination without compensation is less likely to be considered "unfair" or an usurpation of funds the franchisee is entitled to. Although it is just as costly to a terminated franchisee in an opportunity cost sense, when initial lump sum payments have not been made, the franchisor is less likely to have to compensate the franchisee. Therefore, the absence of initial lump sums is crucial for the operation of the self-enforcement mechanism.

³⁵ See, for example, *Zapatha v. Dairy Mart, Inc.*, 381 Mass. 284, 408 N.E. 2d 1370, 1378 (1980) cited in ABA (1990), p. 13, n. 56.

³⁶ See ABA (1990), n. 78, where nine states are listed as having this requirement in their statute.

³⁷ See Klein (1980).

4. Vertical integration

This economic framework has clear implications for when we should see franchisors relying on company ownership as opposed to the franchising of outlets. In particular, condition (2) not only states the determinants of the franchisor's ability to pay a premium to franchisees, but also, equivalently, defines the conditions when the franchisor will find it economically advantageous to vertically integrate, or directly own and operate outlets. The franchisor will own and operate outlets when condition (2) does not hold, or when:

$$W_2 > \sum_t \left[\frac{(C_e - C_f)_t}{(1+r)^t} \right]. \quad (3)$$

When condition (3) is met, i.e., when the present discounted value of the premium stream required to assure franchisee performance is greater than the present discounted value of the cost disadvantage of employee versus franchisee operation, the franchisor will vertically integrate. The franchisor minimizes its total costs by vertically integrating in this case because, in spite of the higher costs associated with employee operation, the franchisor saves the greater costs of having to make a premium payment. Absent contract costs, franchisees could run the system more effectively than employees, but it is too costly for the franchisor to create the necessary contractual relationship with its franchisees under which franchisees would have the incentive to perform. Moreover, even if the franchisor wished to pay the higher required premium stream, if condition (3) holds, the franchisor could not credibly convince franchisees that it would in fact pay the premium.

Much of the empirical work on this question of franchisor ownership of outlets has focused on C_e , i.e., on isolating the conditions where the costs of employee operated outlets presumably are low and, therefore, the incidence of vertical integration can be expected to be high.³⁸ The first pathbreaking empirical work along these lines was done by Brickley and Dark (1987). The primary variable Brickley and Dark find to have a negative effect on franchisor ownership is the distance of an outlet to the nearest franchisor headquarters, which they assume is a reasonable proxy for the cost of monitoring employee-managers in franchisor operated outlets.³⁹ However,

³⁸ Rubin (1978) was the first to argue that the costs of monitoring employee operated franchise outlets was the primary determinant of franchisor ownership of outlets.

³⁹ Brickley and Dark also test Rubin's (1978) suggestion that, because of economies of scale in monitoring units that are more concentrated geographically, monitoring costs are likely to be lower in urban areas than in rural areas. They obtain similar results when they use population of the county that the unit is located in, rather than distance to headquarters, as their proxy for monitoring costs.

while this result certainly expands the systematic information we have about franchising, it does not seem particularly convincing or important. Even if the distance to headquarters positively influences the monitoring costs of franchisor owned outlets, we can expect that such travel costs will also positively influence the monitoring costs of franchised outlets. The crucial variable that should affect the likelihood of vertical integration is the monitoring costs of franchisor owned outlets relative to franchised outlets. Brickley and Dark recognize this but dismiss franchisee monitoring with the assertion that franchisee “quality substitution problems can be reduced by means other than frequent on-site monitoring, e.g., the penalties for cheating can be very severe and can reduce the incentives to cheat”.⁴⁰ However, such a self-enforcement mechanism can and will also be used to reduce monitoring costs of employee managers. A wage premium, for example, may have to be paid to employees when the franchisor vertically integrates in order to get the employees to perform properly and this cost must be included as part of C_e .

While it may be the case that direct franchisor monitoring costs associated with owned and operated units are higher than the direct monitoring costs associated with franchised units, it seems incredible that the difference in direct monitoring costs should be related so strongly to distance to franchisor headquarters and that this distance variable would be so crucial an element in the overall explanation of the degree of vertical integration. The more important considerations would appear to be the determinants of the left hand-side of condition (3), namely the required level of the performance-assuring premium stream that must be paid by the franchisor to franchisees, W_2 , which, in turn, is determined by the extent to which franchisees can deviate in their performance from the desired level, W_1 .

The short-term cheating behavior that can be engaged in by independent franchisees, or W_1 , can be expected to be greater than the short-run cheating behavior that can be engaged in by employees. This is merely “the other side of the coin” of the efficiency gains from franchising. Franchising arrangements are efficient in many circumstances because they harness the efforts of franchisees in building up a business in which they have an ownership right. However, because franchisees are residual claimants they also have an increased incentive compared to employees in deceptively reducing costs and increasing short-run profits in many circumstances. Therefore, the franchisor may choose vertical integration rather than franchising in these circumstances.

Brickley and Dark attempt to measure some of the determinants of short-run franchisee cheating behavior and the relationship of these determinants to franchisor vertical integration, but the results of these much more key tests are, unfortunately, extremely weak. They do find that three indus-

⁴⁰ Brickley and Dark (1987), p. 408.

tries somewhat arbitrarily classified as less likely to have repeat customers (restaurants, hotels and motels, and auto rentals) are more likely to have franchisor owned units. But they also find that units located near freeways, which presumably are less likely to have repeat customers, are more likely to have franchised units.⁴¹

What Brickley and Dark have done is a good first step, a useful form of “feeling the elephant”.⁴² What I would suggest is that further research attempt a more detailed empirical analysis of individual franchise systems in order to more closely examine the incentive incompatibility problems that are present, the ability of transactors to contractually control these problems, the cost disadvantage associated with employee operations, and the importance of these factors in explaining the incidence of vertical integration across systems and across units within a system. What our analysis suggests is that when franchisee malincentives are very large and the franchisor is unable to write a contract directly on desired franchisee behavior or to pay the large required premium, vertical integration may be the cheapest way for the franchisor to assure the supply of desired distribution services. Although employees have a reduced incentive to perform compared to franchisees, vertical integration is the least cost way for the franchisor to control franchisee malincentives.

An example of the type of detailed empirical work I am advocating is the examination by Muris, Scheffman and Spiller (1992, 1993) of the recent movement towards increased vertical integration in the Coke and Pepsi soft drink distribution systems. Muris, Scheffman and Spiller document that Coke and Pepsi’s movement from independent franchised bottlers (with exclusive territorial grants) towards a greater reliance on company-owned bottling operations was due to a number of factors that made the soft drink market environment much more dynamic and complex. One major factor was the growing importance of marketing activity in the industry and the necessity for local bottlers to cooperate in implementing a manufacturer designed marketing campaign by arranging media spot coverage, supplying retailers with adequate product and displays, developing and executing local price promo-

⁴¹ In addition, they find a weak negative correlation across franchise systems between the initial investment that must be made in a unit of a particular system and the percentage of units that are franchised within that system. They explain this on the basis of increased franchisee risk. An alternative explanation may be that as the franchisee’s specific investment increases, the present discounted value of the return on these investments becomes greater than the present discounted value of the cost disadvantage, i.e., the franchisor cannot credibly commit to let the franchisee earn the return on these assets.

⁴² Similar research with similar results has been conducted by Norton (1988), Martin (1988), Brickley, Dark and Weisbach (1991A) and Lafontaine (1992).

tions and, generally, making sure that the product is “pushed” properly, such as making sure retailers supply adequate shelf space.

Unfortunately, Muris, Scheffman and Spiller refer to the difficulties a soft drink manufacturer has in coordinating these tasks with independent bottler franchisees in terms of increased transaction costs. While the narrow contract and negotiation costs they discuss are no doubt high, labeling all the problems generically as transaction costs blurs the important underlying economics. Instead I would emphasize several factors. First, the relatively high price-marginal cost margin that the manufacturer of this differentiated product faces creates an inherent incentive incompatibility problem between the independent franchisee bottlers and the manufacturer. One might be tempted to identify this franchisee malincentive problem as a successive monopoly problem because of the grant to each franchisee of an exclusive territory. However, this mischaracterizes the essential problem. As we have seen, a successive monopoly may be relatively easy to control in many cases and was, in fact, handled by the manufacturer before the market environment changed. Moreover, as we also have seen, a malincentive with regard to the supply of marketing effort exists even when franchisees operate in a perfectly competitive environment.

The second factor I would emphasize is the difficulty of writing direct contracts with franchisees with regard to the supply of local marketing effort. While prices and the successive monopoly problem may be controllable by contract, the non-price marketing services that must be supplied by franchisees are not so easily specifiable or controllable. Further, because the new market environment required franchisees to supply substantially more marketing services, the required franchisee premium is likely to have become substantially higher and, when combined with the declining relative costs of manufacturer operation documented by Muris, Scheffman and Spiller, the manufacturer may not have been able to credibly commit to pay the required premium. Therefore, vertical integration became the most efficient alternative. These considerations are much broader than narrow transaction costs associated with negotiating, writing and revising contracts.

Whether one labels the phenomenon as increased transaction costs or more fully considers the underlying malincentive problems and the difficulties of using both court-enforced and self-enforced contracts to solve these problems, Muris, Scheffman and Spiller correctly contrast their explanation of vertical integration in bottling with an analysis of vertical integration as a response to the hold-up problem associated with firm-specific assets.⁴³ Since there was no obvious increase over time in the extent of specific assets in the bottling industry, the hold-up problem is an unlikely explanation of the

⁴³ See Klein, Crawford and Alchian (1978).

movement over time towards vertical integration. One possibility is to consider the increased marketing activity of the franchised bottlers as in some sense firm specific investments. Because the financial return from marketing activity occurs over time, such activity requires either an explicit long-term contract with the manufacturer to protect these franchisee investments, with the rigidity and other costs associated with long-term contracts, or an implicit contractual understanding with the manufacturer, with the costs associated with an increased premium stream. However, the fundamental incentive incompatibility with regard to franchisee marketing effort that we are focusing on would be present even if there were no firm specific investments and all contracts were short-term. The changing environment substantially increased the manufacturer's costs of monitoring and contractually controlling franchisees and this, along with the required franchisee premium payment, led to vertical integration.

In terms of the more general economic framework of this paper, the hold-up associated with firm-specific assets that I have emphasized in the past should be thought of as only one kind of transactor non-performance, i.e., only one reason why the W_1 remaining after contractual specification may be high. What I have emphasized here is an entirely different kind of non-performance, a kind of non-performance that is not dependent upon the presence of specific investments. And this change in focus creates a different motivation for vertical integration.

Vertical integration works in this situation essentially because of the increased control associated with ownership. Because the franchisor now owns the outlets, the kind of contract the franchisor has with its employees is fundamentally different from the contract it would have with independent franchisee owners. Although the franchisor's employee managers have malincentives because their contract compensates them on the basis of some imperfect measure of performance, the franchisor-employee manager contract reserves all the residual, contractually unspecified rights of the relationship for the franchisor. Therefore, when a franchisor vertically integrates and replaces independent franchisee owners with employee managers, although the possibility for non-performance is not entirely eliminated, key elements of the non-performance problem are eliminated. It is in this sense that the franchisor now has more control over performance.⁴⁴ It is for this added

⁴⁴ For example, in the General Motors-Fisher Body case discussed in Klein, Crawford and Alchian (1978) when General Motors purchased Fisher Body and replaced the Fisher brothers with employees, no employee had the incentive or the ability to impose a cost on General Motors by intentionally mislocating the body producing plant, as the Fisher brothers had done when they were owners of an independent Fisher Body and plant location was an unspecified element of the contract that governed their relationship with General Motors. After vertical integration General Motors substantially increased their control over the operation.

control that franchisors are willing to bear the extra costs of an employee operation when the independent franchisee malincentive problems become severe and the required premium becomes large.

5. Conclusion

Franchise contracts are not unique. The fundamental economic forces that underlie franchise relationships and franchise contracts exist in most distribution arrangements. What is unique about franchising is that it provides us with an accessible source of standardized contracts, many of which have survived for a considerable period of time. These contracts vary across franchise systems and across areas within a franchise system and have evolved over time in response to changes in market conditions. Franchising contracts, therefore, provide us with a good laboratory in which to study contractual arrangements more generally. In addition, because of the basic similarity of the underlying economic forces, the study of franchising is likely to provide us with insights into the economics of distribution and, more generally, with insights into the nature of all vertical relationships.

I have emphasized a number of economic factors in this paper that may serve, in part, as a useful framework in which to conduct our research on distribution contracts. The factors include: the malincentives inherently present in distribution relationships, especially with regard to distributor marketing efforts; the use of contract terms to ameliorate these malincentives by creating a premium stream that facilitates the operation of a self-enforcing mechanism; the costs associated with these contractual arrangements, including the cost to the franchisor or manufacturer of paying the required premium stream; and, finally, the motivation for vertical integration as an efficient solution to the incentive incompatibility problem when these contractual solutions become too expensive.

References

- ABA Antitrust Section: Monograph No. 17, 1990, Franchise Protection: Laws Against Termination and the Establishment of Additional Franchises.
- Brickley, J.A. and F.H. Dark, 1987, The Choice of Organizational Form: The Case of Franchising, *Journal of Financial Economics* 18, 401–20.
- Brickley, J.A., F.H. Dark and M.S. Weisbach, 1991, The Economic Effects of Franchise Termination Laws, *Journal of Law and Economics* 34, 101–32.
- Brickley, J.A., F.H. Dark and M.S. Weisbach, 1991A, An Agency Perspective on Franchising, 20 *Financial Management* 27.
- Dnes, A.W., 1993, A Case-Study Analysis of Franchise Contracts, *Journal of Legal Studies* 22, 367.

- Eckard, E.W., 1985, The Effects of State Automobile Dealer Entry Regulation on New Car Prices, *Economic Inquiry* 24, 223.
- Ippolito, P.M., 1988, Resale Price Maintenance: Economic Evidence from Litigation, Washington, D.C.: Federal Trade Commission, Bureau of Economics Staff Report.
- Kaufman, P.J. and F. Lafontaine, 1994, Costs of Control: The Source of Economic Rents for McDonald's Franchisees, *Journal of Law and Economics* 37, 417–453.
- Klein, B., 1980, Transaction Cost Determinants of 'Unfair' Contractual Arrangements, *American Economics Review* 70, 356–362.
- Klein, B., 1996, Why 'Hold-Ups' Occur: The 'Self-Enforcing Range' of Contractual Relationships, forthcoming *Economic Inquiry*.
- Klein, B., R.G. Crawford and A.A. Alchian, 1978, Vertical Integration, Appropriable Rents and the Competitive Contracting Process, *Journal of Law and Economics* 21, 297–326.
- Klein, B. and K. Murphy, 1988, Vertical Restraints as Contract Enforcement Mechanisms, *Journal of Law and Economics* 31, 265–297.
- Krueger, A.B., 1991, Ownership, Agency and Wages: An Examination of Franchising in the Fast Food Industry, *Quarterly Journal of Economics* 106, 75–101.
- Lafontaine, F., 1992, Agency Theory and Franchising: Some Empirical Results, *Rand Journal of Economics* 23, 263–283.
- Martin, R.E., 1988, Franchising and Risk Management, *American Economics Review* 78, 954.
- Mathewson, F. and R. Winter, 1985, The Economics of Franchise Contracts, *Journal of Law and Economics* 28, 503–26.
- Muris, T., D. Scheffman and P. Spiller, 1992, Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry, *Journal of Economics and Management Strategy* 1, 83.
- Muris, T., D. Scheffman and P. Spiller, 1993, Strategy, Structure and Antitrust in the Carbonated Soft-Drink Industry, Westport, Conn.: Quorum Books.
- Norton, S., 1988, An Empirical Look at Franchising as an Organizational Form, *Journal of Business* 61, 197.
- Rubin, P., 1978, The Theory of the Firm and the Structure of the Franchise Contract, *Journal of Law and Economics* 21, 223.
- Shelton, J.P., 1967, Allocative Efficiency vs. "X-Efficiency", Comment, *American Economic Review* 57, 1252.
- Smith, R., 1982, Franchise Regulation: An Economic Analysis of State Restrictions on Automobile Distribution, *Journal of Law and Economics* 25, 125.
- Telser, L., 1960, Why Should Manufacturers Want Fair Trade? *Journal of Law and Economics*, 86–105.
- Telser, L., 1990, Why Should Manufacturers Want Fair Trade II, *Journal of Law and Economics*.
- U.S. Department of Commerce, 1988, Franchising in the Economy: 1986–1988, Washington, D.C.: U.S. Government Printing Office.
- Williamson, O., 1983, Credible Commitments: Using Hostages to Support Exchange, *American Economics Review* 73, 519.