THE ECONOMICS OF FRANCHISE CONTRACTS*

G. FRANK MATHEWSON and RALPH A. WINTER
University of Toronto

I. INTRODUCTION

As an organizational arrangement, the franchise contract lies between anonymous price-mediated exchange and centralized intrafirm employment. In contrast to conventional market exchange, franchise contracts typically impose on franchisees retail quality standards, common hours of business, price controls, and nonlinear payment schedules (for example, fixed initial royalty fees plus a percentage of gross retail revenues), while franchisors typically provide national advertising and training programs, monitor and inspect the franchisees' performance (with varying intensity across industries), and hold the residual power to terminate the franchise agreement. Observations of actual terminations indicate that franchisors sometimes exclude this power. Franchisors sometimes face persistent queues of potential franchisees. In contrast to simple employment contracts, franchise agreements provide for sharing of profits (through royalties on revenue or output) and considerable freedom for independent decisions by local franchisees.[1] Why do these franchise contracts exist

* We wish to thank Jack Carr, Mukesh Eswaren, Richard Jensen, Yehuda Kotowitz, Ashok Kotwal, Howard Marvel, George Stigler, in particular, and members of seminars at Chicago, McMaster, Ohio State, Southampton, and Toronto for helpful comments; we also thank the Toronto law firm of Goodman and Carr and Harvey Shapiro and Arthur Trebilcock of that firm in particular for the provision of a sample of franchise contracts and Kenneth Fong of McDonald's Restaurants for general discussion on franchising.

[1] The features of franchise contracts appear in Richard E. Caves & William F. Murphy II, Franchising: Firms, Markets and Intangible Assets, 42 S. Econ. J. 572 (1976); Dov Izraeli, Franchising and the Total Distribution System (1972); Alan Karp, Franchising Today: A Specialized Contract (Special Lectures Law Soc'y Upper Canada 1975); Paul H. Rubin, The Theory of the Firm and the Structure of the Franchise Contract, 21 J. Law & Econ. 223 (1978). Franchising accounts for approximately one-third of total retail sales in the United States and Canada. If the central feature of a franchise contract is profit sharing, then our explanation extends beyond narrowly defined franchising to a wide set of vertical contracts, for example, author/publisher, invention/patent licensing.

[*Journal of Law & Economics*, vol. XXVIII (October 1985)]

and how do they resolve agency issues in trade between the parties? What accounts for observed variation in franchise contracts? These are the general issues addressed in this paper.

The principal ingredient in most franchise contracts is the franchisee's right to use a national brand name in exchange for a share of profits to the franchisor. A significant increase in the use of franchise systems occurred in the mid-1950s with an apparent increase in the efficiency of national brand names. This we attribute to three factors. First, the development of television meant that there was a more efficient nationwide information technology, reducing the cost of establishing national brand names. Second, an increase in travel by consumers meant that consumers were more often shopping in unfamiliar geographical areas, in which national but not local brand names would serve as signals of quality, enhancing the value of a national brand name. Finally, a continuing increase in the real income of consumers led to a further increase in the opportunity costs of search in retail markets, again enhancing the information value of brand names.

The use of national brand names by local retailers involves a number of potential agency problems. After a contract has been struck to provide for the retail distribution of a product, the retailer must make decisions on prices, sales effort, and any input (for example, servicing) into the quality of the final product; the manufacturer continues to invest in (national) advertising, product quality, and the product's brand name. In general, if all decisions could be completely specified in the contract, then efficient (joint profit-maximizing) choices could be guaranteed. But costs of enforcing a complete contract, in particular the costs of monitoring the decisions of contractual parties, lead to an incomplete contract. In an incomplete contract, principal's or agent's decisions unspecified in the contract will be undertaken ex post on the basis of the unconstrained self-interest of the decision maker given the incentives provided by the contract; to the extent that this self-interest deviates from the collective interest of all parties to the contract an efficiency (agency) cost is incurred. In a retail contract, in particular, the retailer may "shirk" on his quality input (for example) because he does not capture the entire benefits of an increase in this quality; the manufacturer upstream gains profit, through a positive wholesale markup or revenue sharing as local demand increases (a vertical externality), and to the extent that the retailer's quality input adds to the national brand name of the product both the manufacturer and other retailers benefit. Similarly, the manufacturer has less than full incentive to maintain a strong brand name. If the setup costs of establishing the bulk of the manufacturer's retail distribution system have been sunk by retailers, then the existing retailers must be earning quasi rents (returns on the contract-specific investments), which are

maintained if the brand name is maintained. Retailers therefore share in the benefits of a strong brand name.

In some settings, franchise contracts emerge as an efficient organizational form by minimizing agency (including monitoring) costs, while affording the opportunity for franchisees to employ their knowledge of local markets to produce local services efficiently to the mutual benefit of the franchisor and franchisee. We argue that these vertical agency issues are central to the franchise contract; horizontal free riding by retailers, because of transient consumers, for example, may exacerbate the control problem faced by the franchisor.

The economic literature on general agency issues is extensive. One branch of this literature focuses on observed contracts and practices and, among other things, seeks to explain the separation of ownership and control within the firm. In that work, the firm is viewed as a set of contracts.[2] Another, more formal, branch of analysis seeks to characterize the precise nature of sharing contracts in the presence of stochastic elements in demand or firm costs, costly monitoring (state verification), and limited liability for agents.[3] This paper applies the insights of both of these branches of research.

Section II of this paper sets out the basic model together with necessary and sufficient conditions for a franchise contract with monitoring by the franchisor. Some form of a binding wealth constraint is essential for a profit-sharing franchise contract. A simple wealth-constrained model yields a set of testable propositions.

Section III develops the effect of commitment on franchise contracts and the equilibrium behavior of the franchise and franchisee. For example, franchisees frequently commit by paying a fixed franchise fee at the time of contract signing in return for subsequent quasi rents. Detection of any chiseling means that the contract is terminated and the franchisee forfeits quasi rents. Section III*A* examines the implications that flow from such commitments.

Franchise contracts frequently leave national advertising expenditures

[2] See, for example, Armen A. Alchian & Harold Demsetz, Production, Information Costs and Economic Organization, 62 Am. Econ. Rev. 777 (1972); Michael C. Jensen & William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, 3 J. Financial Econ. 305 (1976); Benjamin Klein, Transaction Cost Determinants of "Unfair" Contractual Arrangements, 70 Am. Econ. Rev. Papers & Proc. 356 (1980); Benjamin Klein, Contracting Costs and Residual Claims: The Separation of Ownership and Control, 26 J. Law & Econ. 367 (1983).

[3] See, for example, Milton Harris & Robert Townsend, Resource Allocation under Asymmetric Information, 49 Econometrica 33 (1981); Bengt Holmstrom, Moral Hazard in Teams, 13 Bell J. Econ. 324 (1982); David E. M. Sappington, Limited Liability Contracts between Principal and Agent, 29 J. Econ. Theory 1 (1983).

and monitoring by the franchisor unspecified in the contract.[4] Rationally designed contracts anticipate the effect of contractual elements on the ex post decisions by the franchisor to advertise and to monitor; that is, advertising and monitoring are self-enhancing. Section III*B* examines the implications for contractual design of the lack of franchisor commitment. Finally, Section IV contains our summary comments and conclusions.

To deal with franchisee free riding, either vertical or horizontal, franchise contracts include vertical restrictions. Frequently, the courts view these restrictions as anticompetitive and disallow them. By providing an efficiency explanation for these contractual restrictions in a simple economic framework, our paper, as a by-product, questions these court decisions.[5]

II. Necessary and Sufficient Conditions for Franchise Contracts

A. Nature of the Problem and Assumptions

We posit a setting where the franchisor or principal (labeled *P*) creates a brand name for a single product through national advertising. The exclusive right to produce and sell this product in a defined territory is leased to a franchisee or agent (labeled *A*); *A* adds additional quality to the product. There are two potential free-riding or externality features to this arrangement: *A* can free ride (vertically) on the national brand name; with transient consumers, any *A* can free ride (horizontally) on the local quality of other *A*'s. With national brand names created and leased by *P*, vertical externalities are always present; horizontal externalities may compound the problem but they are not necessary to explain franchise contracts and restrictions as these arise even with a single *P* and *A*. (Horizontal exter-

[4] Shelby D. Hunt, Franchising: Promises, Problems, Prospects, 53 J. Retailing 77 (1984) cites nondelivery of the promises of franchisors as a problem for franchisees; Dov Izraeli, *supra* note 1, cites franchises where the franchisor failed to advertise as promised after the franchise contract was signed. We examined a (nonrandomly selected) sample of nine franchise contracts in force in Canada, provided by a law firm (Goodman and Carr, Toronto) that writes the majority of franchise contracts in the country. (The contracts were chosen to illustrate the variance in contract restrictions.) In the seven cases where national advertising is provided by the franchisor, brand names in efficiency units were *not* guaranteed. In the eight cases where franchisors monitored franchisees, monitoring included random monitoring at the discretion of the franchisor.

[5] One restriction found contentious in antitrust cases but not dealt with in this paper is exclusive territory. Elsewhere (G. Frank Mathewson & Ralph A. Winter, An Economic Theory of Vertical Restraints, 15 Rand J. Econ. 27 (1984)), we have argued that exclusive territories rationalize price conjectures for competing franchisees. As such, price competition is independent of the incentive issues discussed here; exclusive territories remain outside our model. As a result, we do not deal here with any ex post incentive of the franchisor to violate promises of exclusive territory to franchisees.

nalities are sometimes invoked as the critical source of franchise restrictions.)[6]

Some assumptions are in force throughout the paper and it is useful to specify at the outset these assumptions together with the basic informational asymmetry that is central to the model. The local demand for the franchised product is uncertain. In our model, there are two states of the world: State 1 denotes low-demand or "bad times" and occurs with probability θ_1; state 2 denotes high demand or "good times" and occurs with probability θ_2 ($\theta_1 + \theta_2 = 1$).

The uncertainty of success at any given retail site and the division of the costs of developing the business at any site between P and A seem to be central to most conventional retail franchises. In some franchise systems A may not know prior to signing the contract which retail site will eventually be assigned. Informational asymmetries are highlighted if we assume that subsequent to contract signing and assignment to a territory, A costlessly observes the state of local demand before undertaking any further action; P must rely on A's actions to obtain information on local demand. This type of informational asymmetry is basic to the simplest principal-agent setting and here drives the model.

To focus exclusively on elements of product quality, we assume that price elasticity is constant across states and that price is set once and for all equal to one by P (that is, there is no agency issue on price). The investment P makes in national brand name is defined as Q; A's investment in local quality in state i is defined as q_i.

In the model described to this point, the local retail demand in state i is defined as $X^i \equiv X^i(q_i, Q)$. In addition, we allow the possibility of horizontal externalities across retailers in this model; while these are not necessary to explain profit sharing, their introduction does generate additional testable predictions. We let $\bar{q}$ be the quality set by retailers outside the model and replace q_i in the demand function by $(1 - \alpha)q_i + \alpha\bar{q}$, so that $\alpha(0 \leq \alpha \leq 1)$ measures the potential for local quality unappropriated by the local retailer because of transient consumers. With horizontal externalities, the demand in state i is thus $X^i \equiv X^i[(1 - \alpha)q_i + \alpha\bar{q}, Q]$. We assume that X^i is concave but additively separable in two arguments. That is, changes in national brand name (local quality) do not affect the productivity of local quality (national brand name).[7] For the right to sell the

[6] Rubin, *supra* note 1, at 228.

[7] Additive separability and the state independence of Q in demand eases subsequent comparative statics. In this sense, additive separability can be viewed as a local property. In our model, team production is essential in the sense that sales of the good require strictly positive inputs of local and national quality, that is, $X^i(0, Q) = X^i(q_i, 0) = 0$.

product under the national brand name, A pays P a fee defined to include a fixed component F and a variable component equal to $(1 - f)$ times the realized sales $(0 \leq f \leq 1)$.

The contract thus specifies a payment scheme plus (q_1, q_2), the franchisee quality input in each of the two states, good and bad times. The inclusion of the brand name investment, Q, and monitoring of the franchisee by the franchisor are considered below as well. The possibilities for franchisee chiseling in this contract arise from both vertical and horizontal externalities. As the franchisor cannot (costlessly) identify the realized state of demand, the franchisee has an incentive to free ride by declaring the low state of demand when the high state of demand has occurred, if doing so generates rents to the franchisee. (The franchisee is better informed about local demand than the franchisor.) With output always measured accurately by the franchisor, chiseling by the franchisee takes the form of reducing local quality to free ride on the national brand name and realizing any consequent rents by declaring that state 1 has occurred when in fact state 2 is realized.[8] The consequences of horizontal free-riding are similar. Provided that $\alpha > 0$, each local retailer has an incentive to reduce local quality to the level that yields output consistent with state declaration. In this case, the rents from chiseling again take the form of a saving on local quality.

The quality, $\hat{q}_2$, set in state 2 when A decides to chisel, must satisfy $X^2[(1 - \alpha)\hat{q}_2 + (\alpha\bar{q}, Q)] + X^1(q_1, Q)$, which implicitly defines $\hat{q}_2$ as a function $\hat{q}_2 \equiv \xi(q_1, \alpha, \bar{q})$. This means that the franchisee's quantity signal must be consistent with the declaration of the state of local demand. (The chiseling is independent of Q because of the separability assumption.) We define γ as the proportion by which $\hat{q}_2$ is less than q_1 (see Figure 1) and assume for convenience that γ is independent of q_1 (but is a function of α and $\bar{q}$). Consequently $\xi(q_1, \alpha, \bar{q}) = q_1[1 - \gamma(\alpha, \bar{q})]$ where $\gamma(0, 0) > 0$. This means that, holding constant the magnitude of any horizontal externalities, demand in the two states diverges as local quality increases or equivalently that there is a greater incentive to chisel the larger is the contractual level of local quality in state 1. (These demand curves are illustrated in Figure 1.)

The possibilities for vertical free rides by the franchisor in the contract are straightforward in the case where levels of Q (the national brand name

[8] Here we are invoking the Harris-Townsend result (see Harris & Townsend, *supra* note 3) that an optimal contract with asymmetric information is equivalent to an optimal direct mechanism, that is, a contract in which the informed agent reveals the state to the principal, subject to a constraint that the agent have the incentive, ex post, to reveal the correct state. "Chiseling," in the direct-mechanism interpretation of a contract then refers to misdeclaration of the state.

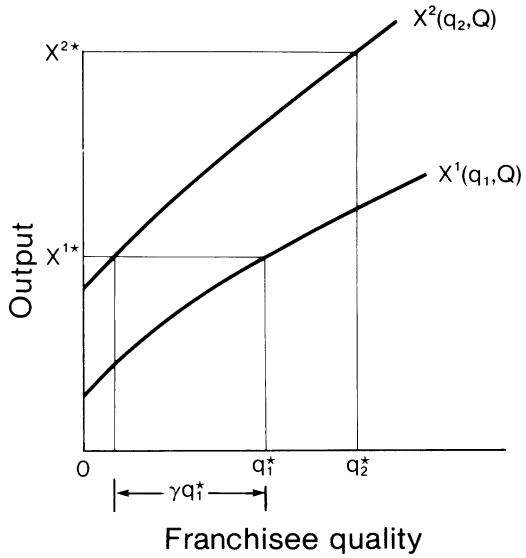

FIGURE 1.—Local quality and free riding

measured in efficiency units) need not be fully enforceable or specified in the contract. For example, when contracts are in effect for twenty years, contingencies affecting Q may be unforeseen; both parties to the contract may be better off when the franchisor has the flexibility to set Q ex post. The freedom to vary Q, however, would leave an incentive for P to reduce Q ex post to the extent that the franchisor does not fully bear the costs of reducing Q because of revenue-sharing arrangements.

The conventional perception is that monitoring is a choice to enforce contracts; in many of the contracts of our sample,[9] the monitoring frequency appears explicitly in the contract. Monitoring as an element of the contract is defined here as ρ, the frequency of local quality verification by the franchisor over a specified time interval. (Monitoring is then similar to crime prevention in the models of crime and punishment.)[10] The costs of quality verification are $C(\rho)$ where $C'(\rho) > 0$. (This monitoring technology is further refined in Section III*B*.) Franchisors are also assumed to observe costlessly output levels and accounting profits for the franchisees. We do assume that these controls are perfect in the sense that retail output levels and low quality levels when monitored are always measured

[9] See note 4 *supra*.

[10] Gary S. Becker, Crime and Prevention: An Economic Approach, 76 J. Pol. Econ. 169 (1968).

accurately (that is, there is no noise in the monitoring mechanism). Finally, franchisors are assumed to incur sunk costs G. These sunk costs are important for our analysis in Section III*A* of fixed franchise fees payable ex ante.

Having set out the specification of our model, we proceed to ask first which contract is first best and then why revenue sharing contracts are used in preference to the first-best contract.

B. First-Best Contracts

With national brand names created ex ante and therefore with the potential for inefficiencies only in franchisee decisions, which contract would elicit optimal local quality decisions and maximum rents for franchisors? The well-known result[11] is that the dominant contract, if feasible, would lease the trademark to a local retailer for a lump-sum payment and establish contractually a monitoring mechanism to detect chiseling or horizontal free riding. A fixed payment ex ante for the local use of the trademark in the presence of an elastic supply of knowledgeable franchisees would maximize the rents to the owner of the brand name; in the absence of horizontal free riding, each retailer would bear the full costs of any reduction in q_i to free ride on the brand name Q; a complete ex ante contracting of monitoring would, by assumption, solve any agency issue for the monitor; each franchisor would have an incentive to seek an efficient complete monitoring mechanism to maximize the rents accruing to the franchisor from the local use of the brand name; in a setting of rational expectations, each franchisee would pay a fee conditional on the local value of the brand name and therefore conditional on the optimal monitoring to reduce horizontal free riding to margins that maximize joint profits given the costs of detection. This shows that an incomplete contract driven solely by an asymmetry on the state of local demand between the franchisor and franchisee is *not* sufficient for profit-sharing franchise contracts.

C. Necessary and Sufficient Conditions for Profit Sharing

In the presence of incomplete contracting to reveal information on local demand, what are the (additional) conditions necessary to generate profit sharing? We argue that with noiseless monitoring some form of a binding wealth constraint for the franchisee is a necessary and sufficient condition

[11] For example, Caves & Murphy, *supra* note 1, at 577.

for a franchise contract.[12] If franchisees were sufficiently wealthy and willing to commit that wealth to guarantee performance, however, sufficient penalties for chiseling would resolve incentive issues for franchisees; variable revenues could accrue exclusively to the franchisor, eliminating any holdup problem. The next section comments on risk aversion; the final section considers contracts under conditions when franchisee wealth constraints are binding. The remainder of this section first develops an argument for limited wealth as the source of franchise contract and then generates a testable implication from this specification.

Consider first a contract where franchisees have zero wealth, an extreme limitation on wealth. This rules out sunk costs by the franchisee. As well, permit contracts to be complete in national advertising and monitoring to eliminate any franchisor holdup. Franchisees have an incentive to free ride on the national brand name—and possibly the local quality of other franchisees. A binding wealth constraint means that franchisors must resolve incentive issues with rewards rather than penalties. With a zero wealth constraint, the franchise contract must generate sufficient surplus in state 1 ("bad times") to pay the royalty fee ex post. Therefore, profits to the franchisee in state 1 are zero. However, profits accruing to A in state 2 ("good times") are necessary to guarantee an incentive for A to reveal truthfully the occurrence of this state. Furthermore, these profits must exceed the expected quasi rents generated by A if state 1 is masqueraded as state 1. Under such a scheme, the royalty fee payable to P in the optimal contract varies with the state of demand, that is, a revenue-sharing scheme is in force. Rents accrue to the franchisee if the expected returns necessary to insure truthful transmission of the local demand states by the franchisee to the franchisor exceed the opportunity cost for the franchisee. In this case, we may observe queues of potential franchisees. (Rents will accrue to franchisees if franchisee wealth is sufficiently low, conditional on the opportunity cost of franchisees.)

D. Risk Aversion

Some analyses of franchising invoke elements of franchisee taste such as risk aversion.[13] In the presence of perfect (nonnoisy) monitoring and nonbinding franchisee wealth, risk aversion by franchisees is *not*

[12] A limited wealth condition is equivalent to a default option on loans to franchisees so that banks incapable of writing performance contracts superior to franchisors will rationally limit their loans to franchisees to facilitate the purchase of the local right to a trademark knowing the incentive issues implicit in the contract.

[13] See, for example, Caves & Murphy, *supra* note 1.

sufficient. Monitoring is critical to the argument. In the absence of monitoring (that is, prohibitively costly monitoring), optimal franchise contracts would levy royalty fees that produce for franchisees losses in state 1 and sufficient surpluses in state 2 so that the contract would both be worthwhile and provide sufficient incentive for franchisees to reveal local demand conditions. (Surpluses from revealing state 2 would have to exceed rents from masquerading state 2 as state 1.) In this case, the optimal franchise contract balances at the margin the provision of insurance by a risk-neutral P (to eliminate profit variability for A across states) with sufficient incentives for truthful state declaration (to leave sufficient profit variability across states for A so that masquerading state 2 as state 1 is unprofitable). Positive (nonnoisy) monitoring, however, in the absence of franchisee wealth constraints permits P to levy a sufficiently large penalty if A is caught misdeclaring the state so that the pay-off to A from truthful state declaration exceeds the expected payoff for misdeclaring state 2 as state 1. The contract then becomes a pure risk-sharing contract with A's income constant in both states. Therefore, risk aversion without a limit on the franchisee's wealth commitment in the presence of monitoring is *not* sufficient for a franchise (profit-sharing) contract. Taste explanations of profit-sharing contracts with perfect monitoring are inappropriate.

Can the sufficient conditions of risk neutrality, unilateral informational asymmetry, and zero franchisee wealth generate testable features of franchise contracts?

E. Franchisees with Zero Wealth: Complete Contracting in National Advertising and Monitoring

From Harris and Raviv's results,[14] we know that any allocation that can be achieved with a contract of royalty fees, local quality monitoring, and national advertising can be achieved with a contract that is incentive compatible and specifies that the agent declares the state. This means that we can restrict our attention to contracts where the agent has no incentive to chisel on local quality and misdeclare the state.

We may now assemble the components to define the optimal contract:

$$\underset{(F,f,q_1,q_2,Q,\rho)}{\text{maximize}}\ F + (1 - f)\Sigma\ \theta_i X^i(q_i, Q) - C(\rho) - Q - G, \quad i = 1, 2, \qquad (1)$$

subject to

$$\Sigma\ \theta_i[fX^i(q_i, Q) - q_i] - F \geq 0, \quad i = 1, 2, \qquad (2)$$

[14] See Milton Harris & Arthur Raviv, Some Results on Incentive Contracts With Applications to Education and Employment, Health Insurance and Law Enforcement, 68 Am. Econ. Rev. 20 (1978).

$$fX^2(q_2, Q) - q_2 - F \geq (1 - \rho)\{fX^2[q_1(1 - \gamma), Q] - q_1(1 - \gamma) - F\} \quad (3)$$

$$X^2[q_1(1 - \gamma), Q] = X^1(q_1, Q), \quad (4)$$

$$fX^i(q_i, Q) - q_i - F \geq 0, \quad i = 1, 2. \quad (5)$$

Equation (1) represents the franchisor's profits net of fixed (sunk) costs; (2) represents the franchisee's profits relative to opportunity costs (assumed to be zero); (3) represents the self-selection (nonchiseling) constraint for the franchisee arising from the asymmetry of information on local demand; (4) reflects perfect auditing of output levels (and can be viewed as defining γ); (5) represents the limited wealth of the franchisee (that is, as royalty fees are paid after production, the contract must guarantee surplus to the franchisee with limited (zero) wealth sufficient to meet the royalty payment). We may substitute (4) into (3) and ignore (2), as (5) means that (2) is always satisfied. Define η and λ_i as the respective shadow prices on (3) and (5).

In state 1 in particular, $fX^1(q_1, Q) - q_1(1 - \gamma) - F > 0$; (4) substituted into (3) means that $\lambda_2 = 0$. Zero wealth means that at the optimal quality and brand name, $F^* = f^*X^1(q_1^*, Q^*) - q_1^*$ (that is, no profits accrue to A in state 1) and therefore, $\lambda_1 > 0$. These two conditions mean that in the light of zero opportunity cost for the franchisee, this contract yields expected rents to the franchisee. Consequently, we expect queues of potential franchisees for these contracts.[15]

These yield the following first-order conditions to define the franchise contract:

$$\lambda_1 = 1 - \rho\eta \quad (6)$$

$$\eta = \theta_2 \quad (7)$$

$$\frac{\partial X^1}{\partial q_1} = 1 + \frac{\gamma\theta_2(1 - \rho)}{\theta_1} \quad (8)$$

$$\frac{\partial X^2}{\partial q_2} = 1 \quad (9)$$

$$\Sigma\, \theta_i \frac{\partial X^i}{\partial Q} = 1 \quad (10)$$

$$\theta_2\gamma q_1 = C'(\rho). \quad (11)$$

[15] Queues of potential franchisees appear to be typical for continuing franchise brand names. Newspaper reports indicate that acceptance rates for franchisee applicants are less than 1 percent for McDonald's and 1.5 percent for Burger King, for example. In a world of heterogeneous labor, queues are a necessary but not sufficient sign of rent. Rents here are different from price premia for honesty in Benjamin Klein & Keith B. Leffler, The Role of

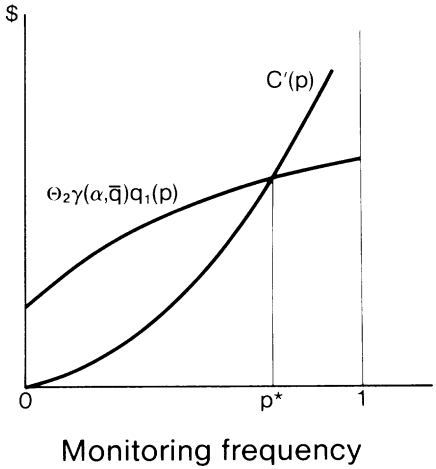

FIGURE 2.—Monitoring equilibrium

These are easily interpreted and reveal several features of the equilibrium contract. These equations indicate that the local quality level in state 2 and the brand name investment are set at their first-best levels (q_2^*, Q^*). In state 1 q_1^* is imposed as a quality floor but set *below* the state 1 first-best quality level in recognition that the rents from chiseling (γq_1^*) are an increasing function of this quality floor. Equations (8) and (11) reveal the nature of the monitoring equilibrium. We may solve (8) for $q_1^* = q_1^*(\rho)$ where $\partial q_1^*/\partial \rho > 0$, so that the marginal benefits of ρ increase in ρ: (8) reveals that q_1^* moves closer to the first-best level as ρ increases. (At $\rho = 1$, equation (8) indicates that q_1^* is set at the first-best level.) A larger ρ, therefore, yields a progressively more efficient setting of q_1^* as monitoring substitutes for a quality floor in the low-demand state artificially lowered as an incentive device. An interior monitoring equilibrium results if $\theta_2 \gamma \cdot \partial q_1/\partial \rho < C''(\rho)$ (provided monitoring is worth the undertaking). (This equilibrium is illustrated in Figure 2.) Should available efficient monitoring technologies yield constant marginal costs, for example, all con-

---

Market Forces in Assuring Contractual Performance, 89 J. Pol. Econ. 615 (1981) where rents are dissipated through sunk expenditures on advertising to yield equilibria (that is, no entry). Rents as a labor incentive device appear in labor market equilibria with involuntary unemployment studied by Carl Shapiro & Joseph E. Stiglitz, Equilibrium Unemployment as a Worker Discipline Device, 74 Am. Econ. Rev. 433 (1984).

tractual arrangements would involve either zero monitoring ($\rho = 0$) or complete monitoring ($\rho = 1$). With $\rho = 0$, contracts would still be profit sharing, with q_1^* set still further below the first-level (some franchise contracts have zero monitoring). With $\rho = 1$, incentive issues in our model disappear as monitoring costs conditions warrant franchisors' becoming fully informed, that is, franchisees disappear from the model as firms are vertically integrated and first-best contracts are achieved. In this case, employees replace franchisees and all profits accrue to the owner of the national brand name.

We may now generate the first comparative static result in our limited-wealth model. Casual empiricism suggests that the variance of local quality within franchised chains itself varies across these chains. How should contractual elements vary in response to observed variation in local quality? It is reasonable to expect this source of local quality variation to flow from the relative marginal cost (productivity) of local quality between realized demand states. It is convenient to respecify the marginal cost of local quality in state 2 as β so that (9) becomes $\partial X^2/\partial q_2 = \beta$. (For example, a limited supply of superior retail sites could alter the marginal costs of local quality creation in state 2.) Comparative statics indicate that $\partial q_2^*/\partial\beta < 0$, $\partial f^*/\partial\beta < 0$ and $\partial F^*/\partial\beta < 0$ while Q^*, q_1^*, and ρ^* remain unchanged. Therefore, driven by this marginal cost effect, the smaller the variance in local qualities (and local demand), the smaller are both the fixed and variable component of the royalty fee schedule. In fact, $\partial[fX^2(q_2, Q) - \beta q_2 - F]/\partial\beta = -q_2 < 0$, so that a fall in the expected rent (and therefore the length of the queue of potential franchisees) is also associated with these reductions in variances.

We could parametrize the marginal cost of national advertising to analyze the effects of changes in the marginal cost of national advertising on these franchise contracts. With our (separable) demand specification, increases in the marginal cost of Q (obviously) lower national brand name size and reduce only the fixed royalty fee, leaving unchanged other contractual elements (including rents accruing to the franchisees in state 2).

This model can also explain the observable practices of franchisors stamping prices on products and enforcing common hours of business. As the contract now stands, nothing prevents A from misdeclaring state 1 when state 2 occurs, providing local quality levels of q_1^* and rationing $X^1(q_1^*, Q^*)$ through another means, such as higher retail prices or shorter store hours. In response to this possibility, P can fix both maximum retail prices by requiring the supplier of the product to stamp the retail price on the good at the factory and by imposing uniform (or minimum) store hours as part of the local quality floor.

III. Extensions and Further Testable Propositions: The Effect of Commitment

In this section, we relax conditions in our model to generate further testable propositions. In each of these models, we focus on a franchise contract with a single representative franchisee in the retail chain. Section III*A* relaxes the strong form of our previous wealth constraint on franchisees (zero wealth). Section III*B* relaxes franchisor commitment by permitting franchise contracts to be specified incompletely in national advertising and monitoring.

A. Franchisee Commitment: Complete Contracting in National Advertising and Monitoring

If we remove the zero-wealth constraint for franchisees, why does the franchise contract not reduce to a first-best contract where franchisees agree to post a performance bond sufficiently large so that with a small amount of monitoring (sufficiently large for a positive probability of detection to be credible) the self-selection constraint is binding for the franchise? In this case, profit sharing would not be required for an efficient contract.

The answer lies in a reverse moral-hazard problem for the franchisor. Monitoring in our model is perfect (nonnoisy). If, however, monitoring were imperfect—if, for example, there were contentious signals on the state of local demand—then the franchisor would have an incentive to misdeclare that the franchisee had breached the contract just to collect the performance bond. Even with perfect monitoring, if the franchisor were to hold a performance bond, the franchisor could be tempted to abscond with the bond and not to deliver the national brand name. The risk is obviously larger where the franchisor has no sunk investment in brand name. Such considerations place limits on any ex ante payment of funds to the franchisor by the franchisee. Symmetrically, if performance bonds are promises by the franchisee to pay large damages to the franchisor in the face of detected breach, rational franchisors anticipate that franchisees who have been detected free riding may disappear (forfeiting any possible future quasi rents) rather than pay large punitive damages.

As an empirical matter, franchisees frequently pay ex ante fixed franchise fees at the time of contract signing. Such fees represent a nonrefundable commitment (or a hostage in Williamson's terms),[16] together with the forfeiture of quasi rents if *A* is caught misdeclaring the state

[16] Oliver Williamson, Credible Commitments: Using Hostages to Support Exchange, 73 Am. Econ. Rev. 519 (1983).

(breaching the contract); this alters the incentive structure for A. What determines the size of such payments? How do they compare to performance bonds? How do they alter franchise contracts?

There is a natural market experiment to assess these questions: new and unproven franchise systems typically carry a low (or even zero) ex ante fixed fee; established and proven franchisors typically demand a more substantial fee. Our objective of generating testable propositions is best served if we analyze the variation in franchise fees and other contractual elements in a simple life-cycle model of new and established franchises.

For this purpose, we add the following assumptions to our basic model: (i) after the execution of the contract including the expenditures on Q, the payoff from Q for a new and unproven franchise can assume one of two values: $h^j(Q)$ with probability μ_j ($j = 1, 2$; $\mu_1 + \mu_2 = 1$), where $\partial h_i/\partial Q > 0$, $\partial h^{j2}/\partial Q^2 < Q$, $h^2(Q^0) \geq h^1(Q^0)$ and $\partial h^2/\partial Q \geq \partial h^1/\partial Q$ (evaluated at Q^0). These conditions are subsequently important for comparative statics: (ii) should outcome $h^1(Q)$ occur, joint profits for P never cover fixed costs G. The franchise venture is therefore bankrupt and does not warrant continuation.

Profits for both P and A are now affected by (exclusive) states on national advertising as well as local demand; for this reason local demands are explicitly written as separable. The difference between the new and established franchise in contract design is simply stated: the new franchise faces the problem of contract design with the value of the brand name given by $\Sigma\ \mu_j h^j(Q)$ and the possibility of bankruptcy after one period; the established (successful) franchise faces the problem of contract design with the value of the brand name given by $h^2(Q)$, because it is known that bankruptcy will not occur. Contracts are renegotiated at the end of the period and for convenience all variables (including sunk costs) are one period. Otherwise, contract design is similar for both firms. To facilitate the definition of the contract, define the (expected) value of the brand name as $\Gamma^k(Q)$ ($k = n$(new), e(established)) where for a new franchise $\Gamma^n(Q) \equiv \Sigma\ \mu_j h^j(Q)$ and for an established franchise $\Gamma^e(Q) \equiv h^2(Q)$.

Will franchisees pay a fixed ex ante fee and what limits the size of this fee or any performance bond? Define $r (0 \leq r \leq 1)$ to be the probability that P or A can successfully abscond (respectively) with any precommitted fee or in lieu of paying any promised damage (r is determined by policing and legal considerations outside our model). Consider first a franchisee precommitment fee of F and then a franchisee performance bond of S. For the moment, fix F for both new and established franchises at some exogenous level F^0 and ask how large F^0 need be to encourage franchisor nonperformance. The delivery of Q by the franchisor is guaran-

teed if $\{(1 - f)[\Sigma\ \theta_i X^i(q_i) + \Gamma^k(Q)] - Q - C(\rho)\} \geq rF^0$ or franchisor quasi rents exceed franchisor expected profits from fraud. (Q and ρ are contractually guaranteed as before if the contract is honored at all by the franchisor.) With a competitive franchise market, expected franchisor profits from performance would be zero or $F^0 + (1 - f)[\Sigma\ \theta_i X^i(q_i) + \Gamma^n(Q)] - Q - C(\rho) - G = 0$ for a new franchise. For an established successful franchise, ex post franchisor profits would be positive or $F^0 + (1 - f)[\Sigma\ \theta_i X^i(q_i) + \Gamma^e(Q)] - Q - C(\rho) - G \equiv \Pi^e(> 0)$. Under competitive entry, the condition that P not abscond may be written as $F^0 \leq G/(1 + r)$ for a new franchise and $F^0 \leq (G + \Pi^e)/(1 + r)$ for an established franchise. This constraint, therefore, is more likely to be binding for a new than for an established franchise because of the higher forgone quasi rent to the successful franchisor from nonperformance; from this constraint alone, profits are more likely to be left with retailers in new than in established franchises. In this case, franchisor incentive issues can be the source of rents with franchisees.

Franchisees in this model will not post any performance bond. Suppose that franchisees not wishing to induce franchisor nonperformance post a bond equal to the difference between the franchisor's expected quasi rents (after payment of sunk costs G) and the expected profits from absconding (that is, $S = (1 - f)[\Sigma\ \theta_i X^i(q_i) + \Gamma^k(Q)] - Q - C(\rho) - rF$). Again, a franchisor incentive problem arises. After delivery of the contractually agreed Q and ρ (and collection of the quasi rents), P would choose $\max\{0, rS\}$ and abscond with (not return) S. (By doing so, P would forfeit any future quasi rent on reputation, Q, but reputation lasts for only one period in this model.)

Symmetrically, any penalty $S \geq 0$ promised by A would not be paid for if A were (accurately) detected chiseling, A would choose $\max\{0, (1 - r)S\}$; therefore, no positive promised S is credible. As a result, performance bonds do not resolve the incentive issues and profit sharing with precommitment of F by P to A where feasible is required. In this case, in general, the contract is designed according to

$$\max_{F,f,q_i,Q,\rho} F + (1 - f)[\Sigma\ \theta_i X^i(q_i) + \Gamma^k(Q)] - Q - C(\rho) - G \tag{12}$$

subject to

$$f[X^2(q_2) + \Gamma^k(Q)] - q_2 \geq (1 - \rho)\{f[X^1(q_1) + \Gamma^k(Q)] - q_1(1 - \gamma)\}, \tag{13}$$

$$\Sigma\ \theta_i[fX^i(q_i) - q_i] + f\Gamma^k(Q) - F \geq 0, \tag{14}$$

$$G - (1 + r)F \geq 0. \tag{15}$$

Equation (12) is the profit relationship for P; (13) is the self-selection constraint for A (F no longer enters this constraint as F is precommitted); (14) reflects A's expected profits—zero in the absence of a binding wealth constraint and an elastic supply of franchisees; (15) reflects in general an incentive for P not to abscond with F under a competitive franchise market. In the following, we assume that (15) is nonbinding. With $F > 0$, the contract design problem is one of maximizing joint profits with respect to the available instruments. (F here, in contrast to our earlier model, has no role in the self-selection constraint but acts solely as a rent transfer device.) First-order conditions defining interior solutions to the contractual elements follow straightforwardly.

Once the contract of the first period is executed, new firms drawing $h^1(Q^*)$ (where * indicates the optimal contractual values) drop out of the market. (In equilibrium failing franchises are μ_1 times the total number of new franchises.) Those remaining know that the market value of their current brand name is $\Gamma^e(Q) = h^2(Q)$. In the subsequent period for established firms, a larger investment in national advertising is warranted and the only uncertainty flows from two states of local demand. While franchisees affiliated with successful brand names realize positive profits in the first period, contracts at the beginning of the second period are renegotiable and franchisors can demand a higher F (provided $F \leq (G + \Pi^e)/(1 + r)$). Other elements of the contract adjust.

The interesting testable issue concerns the difference between the contracts for new and established franchises. This requires some comparative statics. Global comparative statics are difficult given the dimension of the vector of decision variables. We present local results by evaluating the Lagrangian corresponding to each decision variable for the new franchise evaluated at the optimal level for each variable in the established contract.

The levels of q_i are unaltered locally between the contracts (as a result of the separability of local demand in q_i and Q). The comparison of other variables follows:

$$\frac{\partial L^n}{\partial f}(f^{*e}, Q^{*e}) = \eta\rho[\Sigma\, \mu_j h^j(Q^{*e}) - h^2(Q^{*e})] < 0. \tag{16}$$

This expression represents the difference in the forgone quasi rents for a detected free-riding franchisee between a new and an established franchise. As the value of the national brand name is less for a new than for an established firm, the franchisee's temptation to free ride on Q is reduced and f can fall, a franchisee incentive effect. If monitoring were either sufficiently costly to eliminate monitoring ($\rho = 0$) or sufficiently inexpen-

sive to render nonbinding the self-selection constraint for A ($\eta = 0$), then locally the variable franchise fees would be the same across the life cycle of the franchise.

$$\frac{\partial L^n}{\partial Q}(Q^{*e}) = (1 + \eta\rho)[\Sigma\, \mu_j h^{j'}(Q) - h^{2'}(Q)] < 0. \tag{17}$$

The first iteration of the market yields information on the value of the brand names; continuing franchisors recontract to increase national advertising with this information. (This result holds independent of the value of ρ, that is, whether or not $\eta\rho = 0$.)

$$\frac{\partial L^n}{\partial \rho}(Q^{*e}, \rho^{*e}) = \eta[\Sigma\, \mu_j h^j(Q) - h^2(Q)] < 0. \tag{18}$$

Monitoring increases for an established firm in response to the increased value of the brand name and, therefore, the increased temptation for the franchisee to free ride on the brand name. (The result depends on a binding self-selection constraint for A; that is, $\eta > 0$.) The result is consistent with the observed practice in North America of established franchisors' buying back existing franchises to operate outlets with employees directed from head office.[17]

A summary of these (local) comparative static results appear in Table 1.

B. Franchisee Commitment: Incomplete Contracting in National Advertising and Monitoring

Most franchise contracts include some provision for national advertising and monitoring of franchisees, but the specification is usually incomplete. For example, franchisors may establish a national advertising fund but there is no guarantee of specific brand names in efficiency units; franchisors may specify major inspections of retail facilities but undertake unannounced random visits as well. This incompleteness together with

[17] Other reasons advanced to explain the "buy back" include (i) the historically asymmetrical treatment of vertical restrictions by U.S. antitrust authorities—vertical restrictions within firms are unchallenged but between franchisors and franchisees are possibly litigious; (ii) reduced unit monitoring costs with increases in the density of retail networks; (iii) franchisor experience at each retail site means that eventually franchisors learn about local demand conditions and therefore no longer require franchisees. We have two comments. Canadian experience records limited antitrust success at litigating vertical restrictions in general and no intervention in franchise contracts. Yet buy back by established franchisors exists in Canada, largely in metropolitan areas. United States antitrust action may explain the wider margin of buy backs. The Canadian evidence is consistent with ii. All of the evidence is consistent with both iii and our model.

TABLE 1
COMPARISON OF CONTRACTS

	New Franchise (No Commitment) Compared with an Established Franchise (with Commitment)
F	−
f	$-$ $(0 \Leftrightarrow \eta\rho = 0)$*
Q	−
q_i	0
ρ	−

* See note 19 *infra*.

the local informational asymmetry at the heart of our model represents a bilateral asymmetric incentive problem. In this setting, what prevents either party from free riding on the quality inputs of the other, including where relevant, horizontal free riding across the retail network?

Rational franchise contracts should be self-enforcing in the unspecified actions of franchisors. For example, franchisors will make the usual ex post unconstrained marginal decision on national advertising and obviously will free ride on local retailer quality (reduce national advertising) unless franchisors fully bear the costs of this reduction. If franchisors could demand and receive performance bonds from franchisees, then any bilateral incentive problem in our model could again be resolved by jointly (i) requiring the franchisee to post a performance bond, forfeitable on detection of state misdeclaration, and (ii) allowing the franchisor to receive all revenues so that the franchisor fully appropriates the consequences of any advertising reduction (that is, no profit sharing). (This is the standard Alchian-Demsetz resolution of any worker and central contractor incentive compatibility where the worker is dismissed for detected shirking and central contractors keep residual profits.)

We have argued that performance bonds are infeasible as they too create ex post moral-hazard problems. Fixed franchise fees, however, may be feasible. In this setting, profit-sharing (franchise) contracts are again required to resolve the bilateral asymmetric incentive issues. We consider in turn the contracts under incompleteness in national advertising and monitoring; in each setting, P and A are assumed to follow Nash strategies. The results are quite different.[18]

[18] A potential resolution of the bilateral moral hazard problem analyzed in Holmstrom, *supra* note 3, is the introduction of a third agent who performs a role of "balancing a budget." Suppose that a third agent is given ownership rights to the total profit, π (gross of expenditures on quality), in return for offering A and P sharing rates $S_A = -B_P + \pi$ and S_P

Contracts incomplete in national advertising result in franchisors' setting advertising expenditures according to the rule

$$(1 - f)\Gamma^{k'} = 1, \tag{19}$$

so that $Q^k = Q^k(f)$ where $\partial Q^k/\partial f < 0$. (If performance bonds on A were feasible, $f = 0$ would obviously guarantee first-best levels of Q.)

Our distinction between new and established franchises permits us to conclude from (19) that for a constant variable fee (f), the established franchisor would advertise more than the new franchisor (as $\Gamma^{e'} \equiv \partial h^2/\partial Q > \Sigma\, \mu_j \cdot \partial h^j/\partial Q \equiv \Gamma^{n'}$). Optimal contracts are designed knowing $Q^k = Q^k(f)$ so that (19) is self-enforcing. First-order conditions defining the optimal values of the contractual instruments are altered in an obvious fashion.[19]

As with national advertising, the optimal franchise contract requires knowledge of the ex post equilibrium in monitoring by the franchisor and free riding (breach) by the franchisee. For this analysis, we further restrict for convenience the monitoring technology to constant per unit costs H so that $C(\rho) \equiv \rho \cdot H$.

Fixed ex ante franchise fees are permitted up to the level of the expected quasi-rent stream for P (that is, P's nondelivery of Q is ruled out). We define π_l^{*i} as the quasi rents to the franchisor or franchisee ($l = P, A$) gross of monitoring costs. The index $i = 1$ or 2 if A does not free ride under local demand state 1 or 2; $i = c$ if A does free ride under local demand state 2.

$= -B_A + \pi$, respectively. If B_A and B_P are set such that $B_A + B_P = \pi^*$, defined as the total profits when first-best expenditures on quality are undertaken, then each agent, capturing the full residual return, has the "correct" incentive to undertake expenditure on quality and the total outlay of the third party just equals its income (that is, $S_A + S_P = \pi^*$). The problems with this contractual arrangement are that (i) it leaves the two agents with the collective incentive to spend excessively on quality and (ii), in the case of uncertainty, it would leave the third agent with a very risky wealth position. Perhaps for these reasons, the "third-party solution" is not observed in actual franchise contracts.

[19] Incompletely specified national advertising has repercussions for one local comparative static over the life cycle of a franchise. In particular, (17) becomes

$$\frac{\partial L^n}{\partial f}(F^{*e}, f^{*e}, Q^{*e}) = \eta\rho\{\Sigma\, \mu_j h^j[Q^{*e}(f^{*e})] - h^2[Q^{*e}(f^{*e})]\}$$

$$+ (1 + \eta\rho f^{*e})(\Sigma\, \mu_j h^{j'} - h^{2'})\frac{\partial Q^{*e}}{\partial f}.$$

The addition of the second term representing the franchisor's incentive to provide national advertising gives this expression an indeterminate sign. This second term is negative because ex post (self-enforcing) chiseling by P on national advertising is less damaging for a new than an established firm. Therefore, the profit-sharing incentive offered to P $(1 - f)$ can be reduced (f increased). In this case if $\eta\rho = 0$, then from the franchisor's incentive effects alone, f will be larger for a new franchise.

TABLE 2
PAYOFF (Quasi-Rent) MATRIX FOR BILATERAL ASYMMETRIC GAME IN FREE-RIDING AND DELINQUENCY IN MONITORING

A. Local Demand State	Inspect (ρ)	Do Not Inspect ($1 - \rho$)
θ_1	$\pi_A^{*1}, \pi_P^{*1} - H$	π_A^{*1}, π_P^{*1}
θ_2:		
Free ride (δ)	$0, \pi_P^{*2} - H$	π_A^{*2}, π_P^{*c}
Do not free ride ($1 - \delta$)	$\pi_A^{*2}, \pi_P^{*2} - H$	π_A^{*2}, π_P^{*2}

If P monitors A, then P observes the state. If state 2 is drawn and A free rides, then A is penalized, that is, receives zero quasi rent, and P, now knowing the state, receives $\pi_P^{*2} - H$. All other payoffs are straightforward and the payoff (quasi-rent) matrix is shown in Table 2. Inspection of this matrix reveals that, in contrast to a self-enforcing Nash advertising equilibrium outside the contract, monitoring and free riding have no equilibrium in pure strategies. State 2 is the only possibility for free riding by A. In this state, if P chooses to inspect, then A chooses not to free ride; if A chooses not to free ride, then P chooses not to inspect; if P chooses not to inspect, then A chooses to free ride; if A chooses to free ride, then P chooses to inspect, and so on.

A mixed-strategy equilibrium, however, does exist for this monitoring/free-riding game. Define δ to be the probability that A free rides. (We have already defined ρ to be the probability that P inspects the retail site.) Given the expected quasi rents ex post, the respective problems for P and A may be written:

(franchisor)

$$\max_{\rho} E\pi_P \equiv \theta_1\pi_P^{*1} + (1 - \rho)\delta\theta_2\pi_P^{*c} + (1 - \delta(1 - \rho))\theta_2\pi_P^{*2} - Q - \rho H, \tag{20}$$

(franchisee)

$$\max_{\delta} E\pi_A \equiv \theta_1\pi_A^{*1} + \theta_2[\delta(1 - \rho)\pi_A^{*c} + (1 - \delta)\,\pi_A^{*2}]. \tag{21}$$

A solution of these ex post problems yields:

$$\delta^* = H/[\theta_2(\pi_P^{*2} - \pi_P^{*c})] \tag{22}$$

$$\rho^* = 1 - \frac{\pi_A^{*2}}{\pi_A^{*c}}, \tag{23}$$

where

$$1 > \delta^*(f, q_2, Q, H), \quad \rho^*(f, q_2, Q, \alpha, \bar{q}) > 0, \quad \pi_P^{*2} \equiv (1 - f^*) X^2(q_2^*, Q^*) - Q^*,$$

$$\pi_P^{*c} \equiv (1 - f^*)X^1(q_1^*, Q^*) - Q^*, \quad \pi_A^{*2} \equiv f^*X^2(q_2^*, Q^*) - q_2^*,$$

$$\pi_A^{*c} \equiv f^*X^1(q_1^*, Q^*) - q_1^*[1 - \gamma(\alpha, \bar{q})].$$

With this ex post equilibrium, we may now define the ex ante optimal contract as the solution to:

$$\max_{F, f, q_2, Q} F + (1 - f)\{[\theta_1 + \theta_2\delta(1 - \rho)]X^1(q_1, Q) + \theta_2(1 - \delta)X^2(q_2, Q)\} + \theta_2\delta\rho[X^2(q_2, Q) - q_2] - Q - \rho H - G \tag{24}$$

subject to

$$\theta_1[fX^1(q_1, Q) - q_1] + \theta_2\{(1 - \delta)[fX^2(q_2, Q) - q_2]\} + \delta(1 - \rho)\{fX^1(q_1, Q) = q_1[1 - \gamma(\alpha, \bar{q})]\} - F \geq 0, \tag{25}$$

together with (22) and (23).

Again, we assume that values of F that transfer rents to the franchisor are not sufficient to deter franchisor contract performance. First-order conditions that define optimal contractual values are once again straightforward. From (22) and (23), $\partial\delta^*/\partial f = H[X^2(q_2^*, Q^*) - X^1(q_1^*, Q^*)]/[\theta_2(\pi_P^{*2} - \pi_P^{*c})^2] > 0$ and $\partial\rho^*/\partial f = -[\pi_A^{*c}X^2(q_2^*, Q^*) - \pi_A^{*2}X^1(q_1^*, Q^*)]/(\pi_A^{*c})^2 < 0$, so that increases in the franchisee's share of the variable profits in equilibrium increase the probability that franchisees free ride and decrease the probability that franchisors monitor. The first-order condition for f is:

$$-H\frac{\partial\rho}{\partial f} = -\theta_2\{[X^1(q_1, Q) - q_1][1 - \gamma(\alpha, \bar{q})] - [X^2(q_2, Q) - q^2]\}\left[(1 - \rho)\frac{\partial\delta}{\partial f} - \delta\frac{\partial\rho}{\partial f}\right] \tag{26}$$

which has the necessary signs to define $0 < f^* < 1$. Therefore, the optimal ex ante contract when monitoring is incompletely specified has an interior equilibrium on f that, in turn, yields a positive probability that simultaneously the franchisee will free ride and the franchisor will inspect, detect the breach, and terminate the contract. This result stands in contrast to the other contractual results in this paper where the contract is designed to prevent breach. Therefore, breach and its detection are an equilibrium result of the model, not an ad hoc external phenomenon. With perfect

monitoring, breach and its detection do not result in litigation; litigation would have to flow from imperfect monitoring.

IV. Conclusions

The objective established for this paper at the outset was an explanation of the use of franchise contract (revenue or profit sharing) to resolve agency issues and the generation of testable propositions to account for the variability of observed franchise contracts. Incompleteness in contracts (for example, in the revelation of local stochastic demand) with team production of quality is *not* sufficient for franchise contracts. However, some form of binding franchisee wealth or limitations on commitment and the incompleteness of contracts are both necessary and sufficient for franchise contracts. Vertical externalities are essential for franchises; horizontal externalities are not. Franchise contracts need not depend on elements of taste, such as risk aversion, but can be generated from incentive considerations alone.

The predictions from our model follow:

i) Conditional on the marginal cost of monitoring, the probability of monitoring is in equilibrium between zero and one.

ii) Franchises in equilibrium may earn positive profits (rents) as incentive devices with resulting queues of franchisees as a consequence of a binding franchisee wealth constraint.

iii) Royalty fees are nonlinear and franchisors impose minimum quality standards (floors) on franchisees; further, franchisors have an incentive to affix prices to the products and regulate the franchisees' hours of business.

iv) If we compare franchise chains, a smaller variance in the local quality and demand at the retail outlets in the system should be associated with a smaller variance in royalty fees (both the fixed and variable fees are reduced).

v) With complete contracting in national advertising and monitoring and ex ante fixed franchise fees, new franchise chains with uncertain brand names relative to established franchise chains with successful brand names, ceteris paribus, should have lower franchise fees (both fixed and variable), lower national advertising, less monitoring, but identical local quality levels.

vi) With contracts incomplete but self-enforcing in national advertising, a Nash advertising equilibrium has less national advertising than an equilibrium with committed national advertising.

vii) With contracts incomplete but self-enforcing in monitoring, the only ex post Nash equilibrium in free riding and monitoring delinquency is

in mixed strategies; in contrast to the other franchise contracts in this paper, which are separating in the sense that contract breach is *never* observed, contracts incomplete in monitoring yield, in equilibrium, contract breach and detection and therefore franchisee termination.

Bibliography

Alchian, A., and Demsetz, H. "Production, Information Costs and Economic Organization." *American Economic Review* 62 (1972): 777–95.

Becker, G. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (1968): 169–217.

Caves, R., and Murphy, W. "Franchising: Firms, Markets and Intangible Assets." *Southern Economic Journal* 42 (1976): 572–86.

Harris, M., and Raviv, A. "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance and Law Enforcement." *American Economic Review* 68 (1978): 20–30.

Harris, M., and Townsend, R. "Resource Allocation under Asymmetric Information." *Econometrica* 49 (1981): 33–64.

Holmstrom, B. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (1982): 324–40.

Hunt, S. "Franchising: Promises, Problems, Prospects." *Journal of Retailing* 53 (1977): 71–84.

Izraeli, D. *Franchising and the Total Distribution System*. London: Longman, 1972.

Jensen, M., and Meckling, W. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (1976): 305–60.

Karp, A. "Franchising Today: A Specialized Contract." Special Lectures. Toronto: Law Society of Upper Canada, 1975.

Klein, B. "Transaction Cost Determinants of 'Unfair' Contractual Arrangements." *American Economic Review Papers and Proceedings* 70 (1980): 356–62.

Klein, B. "Contracting Costs and Residual Claims: The Separation of Ownership and Control." *Journal of Law and Economics* 26 (1983): 367–74.

Klein, B., and Leffler, K. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 89 (1981): 615–41.

Mathewson, F., and Winter R. "An Economic Theory of Vertical Restraints." *Rand Journal of Economics* 15 (1984): 27–38.

Rubin, P. "The Theory of the Firm and the Structure of the Franchise Contract." *Journal of Law and Economics* 21 (1978): 223–34.

Sappington, D. "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory* 29 (1983): 1–21.

Shapiro, C., and Stiglitz, J. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74 (1984): 43–44.

Williamson, O. "Credible Commitments: Using Hostages to Support Exchange." *American Economic Review* 73 (1983): 519–40.