Dual Distribution in Franchising as an Efficient Monitoring Device under Asymmetric Information*

Thierry PENARD
CREM, University of Rennes I
Faculté des Sciences Economiques, 7 Place Hoche,
35065, Rennes cedex, France
e-mail: thierry.penard@univ-rennes1.fr

Emmanuel RAYNAUD
INRA SADAPT & Centre ATOM (University of Paris I)
INRA SADAPT, 16 rue Claude Bernard,
75231, Paris cedex 05, France
e-mail: eraynaud@inapg.inra.fr

Stéphane SAUSSIER (corresponding author)
ADIS, University of Paris XI & Centre ATOM, University of Paris I,
Maison des Sciences Economiques, 106-112 Bd de l'Hôpital,
75013, Paris, France
e-mail: saussier@univ-paris1.fr

June 2006

* We would like to thank one of the coeditors and the referee for their very useful contribution to significantly improving the paper. Gratitude also gets addressed to James Brickley, Eric Brousseau, Michael Raith, Joel Ruet, Vivian Dos Santos Silva and Josef Windsperger for their helpful comments. A previous version of this paper was presented at ATOM (University Paris I), INRA ESR (Dijon), the ISNIE conference held in Boston (MIT), the EMNET conference in Vienna and at the fourth IIOC Conference in Boston (Northeastern University).

ABSTRACT:

One recurring theme that emerges from empirical studies on franchising is the coexistence of franchised and company-owned units within the same chain. This paper supports the idea that mixed chains or dual distributions correspond to efficient organization when both the behavior of managers and the nature of local markets are costly to observe. We will start by showing that partial monitoring, i.e. when the franchiser only monitors a subset of his outlets, represents an optimal strategy. We will then prove that the coexistence of franchised and company-owned units economizes on monitoring costs and can be analyzed as an efficient – and non-transitory – organizational choice.

KEYWORDS: dual distribution, monitoring, asymmetric information

JEL classification: D23, L14, L22

1. INTRODUCTION

Franchising has been receiving considerable attention in the empirical and theoretical literature on contracting. Most works have focused either on the determinants of contractual provisions, especially royalty rates (Mathewson and Winter, 1985; Lafontaine 1992; Lafontaine and Shaw, 1999; Scott, 1995; Brickley, 2002), or on the tradeoff between franchising and company ownership (Brickley and Dark, 1987; Lafontaine, 1992; Brickley, 1999).

One consistently puzzling empirical finding mentioned in many empirical studies is the coexistence of franchised and company-owned units within chains (sometimes called the "plural form", "contract mix" or "dual distribution"). Most of the literature views this phenomenon as *transitory*. For those who argue that franchising is more profitable than company ownership, a franchiser may initially operate several units either to demonstrate the quality of its business plan to potential franchisees (Gallini and Lutz, 1992; Lafontaine, 1993 for an empirical assessment) or to credibly commit to protecting the value of its brand name (Scott, 1995). The extent of company ownership however is likely to decrease with chain maturity and brand name reputation. Other authors assert that the rationale for franchising should instead disappear with chain maturity (Oxenfeldt and Kelly, 1969). Franchising yields temporary access to certain scarce resources, either capital (Caves and Murphy, 1976), managerial talent (Norton, 1988) or local information (Minkler, 1990), which facilitate expansion. As firms become established, they should reduce their reliance upon franchising (ownership redirection).

These two extreme franchising evolution patterns (i.e. franchising redirection, ownership redirection) have not been empirically validated. Using a large panel data set from

US and Canadians chains, Lafontaine and Shaw (2005) have shown stability in the franchise mix ratios as chains mature. Similar evidence has been observed in other countries (see Furquim de Azevedo and dos Santos Silva, 2001 for Brazil or Pénard, Raynaud and Saussier, 2003 for France). This would suggest that dual distribution may be an *efficient and persistent* organizational form.

Several explanations have been forwarded for this steady-state form. Bradach (1997) emphasized the complementarities between these two contractual arrangements in order to maintain quality and homogeneity throughout the units while at the same time promoting innovation. He argued that "*chain organizations are more than the sum of their parts: by having both company and franchise arrangements together, a chain can leverage some of the strengths and overcome some of the weakness associated with each arrangement*" (1997, p. 279). Lewin-Solomon (1999) formalized part of Bradach's explanation by justifying the existence of dual distribution as a commitment device used by a franchiser to incite franchisees to innovate. In most chains, franchisees pay a royalty rate that equals a fraction of their revenue (and not their profit). With such a royalty structure, chains will favor innovations that increase revenues, sometimes at the expense of franchisee profits. By owning some units in the chain, the franchiser's interests are better aligned with those of his franchisees. An alternative explanation was provided by Bai and Tao (2000), who considered a multitask model along the lines of Holmström and Milgrom (1991) and showed that dual distribution is an optimal means for inducing effort towards both sales and brand name development. Furthermore, Sorenson and Sorensen (2001) explained franchise mix as an efficient learning organization, resulting from a tradeoff between exploration (franchising) and exploitation (company-owned units). Over the course of these explanations, dual distribution would appear to be an efficient and stable arrangement.

This article provides another rationale for dual distribution. It has long been argued that moral hazard and related monitoring costs are important drivers of chain organization, yet mainly used to explain contract design (Lafontaine, 1992; Mathewson and Winter, 1985; Rubin, 1978) and the "make or buy" decision (Brickley and Dark, 1987; Lafontaine and Slade, 1996). Brickley and Dark (1987), for instance, predicted that as monitoring costs (i.e. direct supervision costs) rise, chains will rely more heavily on franchising. The premise of this article is that moral hazard and monitoring costs also constitute key factors in explaining the prevalence of "make *and* buy". Dual distribution can indeed prove to be a monitoring cost-saving strategy for mitigating managers' moral hazard.

For this purpose, we have developed a model in which a chain needs to monitor the managers of its outlets under conditions of asymmetric information on local demands and managers' efforts. The outlets are located in various commercial areas and subject to a random fluctuating demand. Since the chain imperfectly observes local demand shock and manager behavior, two kinds of moral hazard appear depending on the manager's status. In a company-owned unit, the salaried manager is paid a fixed wage and therefore tempted to shirk, by making less of a sales effort than expected. In contrast, within a franchised unit, the franchisee is the residual claimant and incited to deploy an intense sales effort; here, the franchisee's moral hazard assumes the form of declaring less than actual sales. A franchisee can announce a sales level below actual sales in order to extract rent ("royalty evasion"), which leads to a sharp contrast with the previous literature on franchising, which implicitly assumes that individual sales (or outputs) are observable at no cost to the chain (see however Mathewson and Winter, 1985 and Gal-Or, 1995 for exceptions). In our model, the chains must incur a strictly positive cost to assess the actual sales of individual units. From such a perspective, this cost can also be viewed as the expense to the chain to make sales or revenues

verifiable.[1] Bradach's work (1997) on the restaurant industry partially supports this assumption by claiming that while headquarter management information systems exist for certain company units, such is not the case with franchised units: more specifically, "*the only financial information that the chain received was a revenue number each month from which the royalty was calculated*" (1997, p. 288). Combs *et al.* (2004) also cite this cheating hazard as one way for a franchisee to hurt the chain[2].

This problem of managers' moral hazard (either under-declaration of sales or shirking) can be circumvented by monitoring the units, for instance through direct supervision or auditing. However, monitoring the units is costly and this cost increases with the number of monitored units (or with the intensity of monitoring an individual unit). It might therefore be sub-optimal to monitor all units (or to monitor all of them with the same intensity), as extensively shown in the literature on auditing within the context of tax evasion, utility regulation and antitrust enforcement (Baron and Besanko, 1984; Mookherjee and Png, 1989; Khalil, 1997). Among this body of literature, the model derived by Besanko and Spulber (1989) deserves special attention and provides a fruitful framework for addressing this issue. The authors characterized the optimal monitoring policy of an antitrust authority that imperfectly observes firms' production costs and pricing behavior. They proved that it

[1] Most of the literature on franchising, and more generally on contract theory, distinguishes observable variables from unverifiable variables by assuming that the former are observed at zero cost, whereas the latter are too costly to observe. We might argue in the real world that verifiability is an agent's decision based on a cost/benefit analysis. Some variables are costly to observe and verify, but this does not necessarily deter parties from contracting on them (Khalil and Lawarée, 1995). Recent literature has begun to explicitly introduce the notion of verifiability as a decision based on the cost and incentives to produce evidence for a third-party enforcer (see for instance Bull and Watson, 2004).

[2] The other hazards respectively are: (i) not adhering to quality standards, and (ii) divulging the chain's proprietary information.

may be optimal to tolerate a limited amount of collusion ((*i.e*. to adopt a *laissez-faire* policy for slightly harmful price agreements[3]). The antitrust agency then simply weighs the costs vs. welfare gains of this audit.

We are interested here in determining the optimal monitoring policy for a chain unable to observe the behavior of its units' managers. We will first consider the case of a "pure" chain (either wholly franchised or company-owned) and prove that complete monitoring is never optimal. Chains should relax inspections on those units declaring or reaching high sales revenue since offering them rent represents a cost-saving method for the chain that deters engaging in more opportunistic behavior (under-declaring sales or shirking). Given this monitoring policy, we will also demonstrate that dual distribution is more profitable than a pure chain. A plural form enables reducing asymmetric information on local demand and curtailing the extent of opportunistic behavior. Salaried managers exert a sort of pressure on franchisees, preventing them from declaring very low sales; this same pressure is placed on the salaried personnel, preventing them from limiting their efforts. Through dual distribution, the franchiser is able to save on the rents that must be awarded to the most efficient units. Moreover, we will show that royalty rates and the extent of company-owned units are positively correlated and appear to be complementary tools in the overall organizational design of the chain.

This article thus provides added support to the commonly-held idea that the coexistence of franchised and company-owned units relies on the heterogeneity of outlets, including distance from the headquarters (Brickley and Dark, 1987)[4], percentage of repeat

[3] The optimal policy calls for an industry audit when the observed market price is too high (i.e. above a threshold price) and for no audit when the market price is reasonable (under the threshold price).

[4] Brickley and Dark (1987) used geographical distance as a proxy for direct monitoring costs and showed that outlets far from the franchiser's headquarters tend to be franchised. When monitoring management behavior

business (Brickley, 1999). The rationale here for franchise mix stems from the chain's inability to observe at no cost the demand levels experienced by each unit. Our paper can also be related to those by Mathewson and Winter (1985) and Gal-Or (1995), both of which proposed a setting wherein managers can take advantage of information asymmetry at the actual level of local demand. Mathewson and Winter (1985) however are primarily concerned with the efficiency of sharing contracts within the context of a single franchiser-franchisee pair and does not deal with dual distribution. Using a similar setting, Gal-Or (1995) found that it sometimes is in the chain's interest to monitor just a subset of outlets. In the Gal-Or paper, the extent of monitoring, i.e. monitoring some or all outlets, is equivalent to the extent of vertical integration since it is assumed that monitoring a unit will allow the chain to dictate the desired level of effort. In our model however, monitoring is not equivalent to ownership. Some units are monitored while remaining franchised and others are company-owned but not monitored. The proportion of monitored units within the chain thus differs from the extent of vertical integration.

This paper will proceed as follows. Section 2 will present the theoretical model, while the set of propositions will be displayed and discussed in Section 3. Section 4 provides some concluding remarks.

2. THE MODEL

2.1 Model framework

The model considers a chain with outlets located in different geographical areas. Each local market is identically characterized by a random fluctuating demand. The positive or

---

proves costly, chains rely more on the incentive provided by a franchise contract to reduce managerial incentives to shirk (see also Norton, 1988, Lafontaine, 1992, Bercovitz, 2001, and Lafontaine and Slade 2001 for survey data).

negative shocks affecting local demand are observed by store managers and not by the chain owner. Moreover, random shocks on a local market are independent of what happens on the other local markets[5]. Let θ be the demand level experienced by an outlet, where θ is distributed between $\underline{\theta}$ and $\overline{\theta}$ according to a pdf f and cdf F. A negative shock tends to push demand downwards (towards $\underline{\theta}$) while a positive shock increases demand towards $\overline{\theta}$.

Each outlet is run by a manager who can either be a franchisee or an employee. All managers are assumed to have the same abilities in running a store. Store performance will depend however on the level of demand. Since managers are perfectly informed on local market conditions, the chain must thus rely on them knowing the true level of local demand.

We now define $V_F(\theta)$ as the expected revenue of a store experiencing a level of demand θ, with $V_F'(\theta) > 0$ and $V_F''(\theta) < 0$ [6]. For a store with a demand level of θ however, the actual revenue can be less than $V_F(\theta)$ because sales also depend on the efforts expended by the manager. For the sake of simplicity, these efforts e assume a value of between 0 and 1, and the cost of efforts is given by δe. Managers' efforts exert a positive impact on sales and, once again for simplicity, the actual revenue of a store experiencing demand θ and run by a manager deploying effort e is defined by: $eV_F(\theta)$. A manager who expends no effort will achieve zero sales (regardless of the level of demand), whereas the same manager expending the highest level of effort (e=1) will generate the maximum revenue $V_F(\theta)$ [7]. Moreover, the

[5] This rules out the possibility of comparing information across markets and implements a tournament-like type of contract in order to mitigate uncertainty.

[6] Independently of the level of demand, $V_F(.)$ should be positively influenced by the value and reputation of the chain ("brand name effect").

[7] We will see later on that the chain is always able to enforce an effort level above zero with a salaried manager, ruling out the possibility of actually observing units with no sales.

current operating costs of a store, net of manager earnings, are constant and equal to C, regardless the store's status.

We will now define both manager and chain payoffs depending on the status of the outlet (whether franchised or company-owned).

2.2 Payoffs within a franchised unit

For all franchised units, we have assumed an identical contract involving the payment of sale-based royalties β[8], which is consistent with franchiser practices: in reality, franchisee contract customization is rarely observed (see Bhattacharyya and Lafontaine, 1995; Lafontaine, 1992). Franchisee profit on a market with demand θ is thus given by $(1-\beta)eV_F(\theta)-C-\delta e$, while the franchiser receives a revenue of $\beta eV_F(\theta)$. Franchisees can potentially display two types of opportunistic behavior. On the one hand, they can shirk and put forth limited effort[9], while on the other, a franchisee can under-declare the actual value of sales revenue and save on the amount of royalties paid to the chain (a behavior similar to tax evasion). This form of opportunism could be costly to detect since a low declared sales revenue figure can result from an honest franchisee experiencing low demand or from an opportunistic franchisee experiencing high demand. In order to better focus on this type of opportunistic behavior, we have ruled out potential shirking by franchisees through assuming:

Assumption 1: $(1-\beta)V_F(\underline{\theta})>\delta$ (1)

The marginal benefit of additional sales effort to a franchisee is always assumed to be higher than the marginal cost of his effort. The franchisee will thus always expend the

[8] For the sake of simplicity, we have assumed that the franchiser does not claim initial fees from his franchisees.

[9] The severity of this moral hazard problem increases with royalty rates (Lafontaine, 1992).

maximum level of effort (e=1)[10]. We will now further define the cost of monitoring franchisees' sales and discuss how franchisees react to monitoring.

2.3 Payoffs within a company-owned unit

Managers in company-owned units receive a fixed wage w regardless of the level of effort they put forth. We consider herein that the counterpart of this wage is to exert the maximum effort (e=1). For a demand level θ and manager effort e, the chain's expected profit is $eV_F(\theta) - w - C$, and the manager's utility is $w - \delta e$. We will now assume that all outlets are profitable regardless of demand level, provided the manager is exerting maximum effort.

Assumption 2: $V_F(\underline{\theta}) - w - C > 0$ (2)

Since the salaried manager payoff decreases with sales efforts expended, the manager has a strong incentive to shirk. The chain is not able to directly observe the manager's efforts, but instead can use the sales or revenues from its outlets to infer the effort being made. As an example, if sales of below $V_F(\underline{\theta}) > 0$ were recorded, it would be known with certainty that the manager had been dishonest (exerting effort beneath what was required) and he could be punished. The chain can therefore always enforce a minimum, yet strictly positive, level of effort. Regardless of demand, it remains in the interest of salaried managers to select an effort e such that sales exceed $V_F(\underline{\theta}) > 0$. In accordance with Assumption 2, the franchiser will be guaranteed positive profits with this minimum sales level.

[10] We have ruled out this type of opportunistic behavior from our model since this problem is less prominent than the under-declaration hazard. If we relax assumption 1, the chain would provide incentives both to mitigate shirking and under-declaration; moreover, it will significantly complicate the model without adding any insight. One way to reinforce assumption 1 is to suppose that franchisees' efforts also reduce the cost of sales, thereby lowering franchisee incentive to shirk.

2.4 Monitoring policy and moral hazard

To overcome the moral hazard stemming from managers (both franchisees and salaried personnel), the chain may proceed by monitoring them (e.g. auditing their accounts and/or conducting visits)[11]. We have assumed that monitoring technology is "perfect", yet remains costly, *i.e*. an audit enables the franchiser to determine with certainty the actual level of demand and consequently to know whether the manager has been cheating (see Gal-Or, 1995 for a similar assumption). The cost of auditing an outlet is fixed and denoted by K if the store is company-owned and Q if the store is franchised. We can presume that Q<K since sales are likely to be easier to monitor (output monitoring) than effort (input monitoring), yet Q is always positive even if only slightly[12].

We will now describe an efficient monitoring policy for the chain. Let's start by noting that it is always better to design a monitoring policy that takes advantage of available information rather than opting for a myopic or random policy. In the present case, the existing information is franchisees' declared sales and observed sales from company-owned units. Through such sales, the manager is more or less stating the level of demand. When a franchisee experiencing demand θ declares a revenue R to the franchiser, this is therefore equivalent to announcing a demand θ_R with $\theta_R = V_F^{-1}(R)$, where $V_F^{-1}(.)$ is the inverse revenue function[13]. If the franchisee is behaving honestly, then $\theta_R = \theta$, whereas and if he is behaving opportunistically (*i.e*. $R < V_F(\theta)$), then $\theta_R < \theta$. In any event, the franchiser's

[11] In this instance, our understanding of monitoring encompasses all monitoring activity to accurately measure sales and assess behavior (Lafontaine and Slade, 1996).

[12] Q may be close to zero for many types of activities (product sales) yet remain high for other activities (services), for which sales monitoring is more complex and costly, e.g. in a fast-food franchise, the franchisee can declare that a certain number of hamburgers were discarded when, in reality, they were sold.

[13] Because the franchiser knows that the franchisee's maximum effort is (e=1).

revenues are given by: $\beta V_F(\theta_R)$. Likewise, when a salaried manager experiencing demand θ earns revenue *R*, it is equivalent to announcing demand θ_E, with $\theta_E = V_F^{-1}(R)$. If the salaried manager is behaving honestly (*e*=1), this yields $\theta_E = \theta$ and $\theta_E < \theta$ otherwise.

The optimal monitoring policy will thus be conditional upon the declared or announced demand (θ_R and θ_E). Let $\gamma(\theta_R)$ denote the probability of monitoring a franchisee who declares sales of $V_F(\theta_R)$ (*i.e*. announces a demand of θ_R) and let $\varpi(\theta_E)$ represent the probability of monitoring a salaried manager who generates sales of $V_F(\theta_E)$ (*i.e*. announces a demand of θ_E). If $\gamma(\theta_R)$=1, then the franchisee knows with certainty that he will be monitored and if $\gamma(\theta_R)$=0, no monitoring will occur. In the intermediate case, the franchisee will be inspected with probability or frequency $\gamma(\theta_R)$ [14]. The same applies for $\varpi(\theta_E)$.

The monitoring policy must also specify punishment schemes. In company-owned units, if the chain detects an effort of below 1, it imposes a penalty *b* on the salaried manager[15]. Similarly, if the franchiser detects under-declaration of revenues, he penalizes by *a* the franchisee either by terminating the contract or through introducing a fine[16]. In both cases, the penalties are assumed fixed and independent of the damage[17]. Moreover, it has been

---

[14] An alternative interpretation would be to consider that over period T, the franchisee receives $\gamma(\theta_R)$ T visits.

[15] These punishments can assume different forms, ranging from a fine to dismissal. The termination of a labor contract, for example, represents a credible punishment if the manager is paid above the market wage, at an efficiency wage level (Shapiro and Stiglitz, 1984) or if the manager has invested in specific human capital, whether partly or entirely financed by him. See Krueger (1991) for empirical evidence of the efficiency wage for employees within company-owned units in the fast food industry.

[16] See Klein (1995) and Lafontaine and Raynaud (2002) for insights on self-enforcement in franchising where contract termination (coupled with expected rents) acts as an incentive device.

[17] *a* and *b* can also denote maximal manager liability. If agents are risk neutral, then the optimal monitoring

assumed that:

$$\textit{Assumption 3: } a \geq \beta\left[V_F(\bar{\theta}) - V_F(\underline{\theta})\right] \tag{3}$$

$$\textit{Assumption 4: } b \geq \frac{\delta\left(V_F(\bar{\theta}) - V_F(\underline{\theta})\right)}{V_F(\bar{\theta})} \tag{4}$$

Assumptions 3 and 4 indicate that if a manager is certain to be monitored ($\chi(\theta_R)=1$ and $\varpi(\theta_E)=1$), the penalties are sufficiently high to deter cheating, regardless of the level of demand. The left-hand side represents the cost of cheating and the right-hand side the highest benefit gained from cheating (*i.e.* that derived by the manager experiencing the most favorable demand conditions while declaring the lowest demand).

Figure 1 displays the timing sequence. In Step 1, the franchiser organizes the chain (opting either for a pure system or a plural form) and announces his monitoring policy. During Step 2 and depending on the monitoring policy, the franchisee declares a sales level $V_F(\theta_R)$, and the salaried manager chooses a level of effort that induces sales of $V_F(\theta_E)$. In the last step, the chain implements its monitoring program.

---

policy involves a penalty at its highest level in order to reduce the frequency of controls (which by nature are costly) – see Becker (1968).

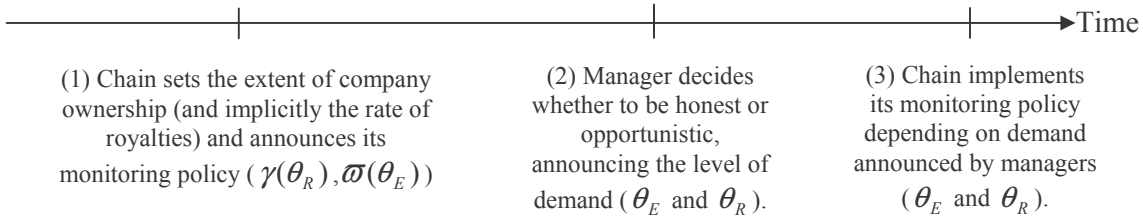

Figure 1: The timing of decisions

In the next section, we will solve the strategic game and determine the optimal monitoring policy under different organizational forms.

3. THE EXTENT OF MONITORING AND DUAL DISTRIBUTION OPTIMALITY

3.1 Pure systems and complete monitoring

A purely franchised system

Let's first consider the case of a purely franchised chain. If the franchiser is seeking to dissuade a franchisee experiencing demand θ from declaring $\theta_R < \theta$, then he must announce a monitoring probability $\gamma(\theta_R)$, such that:

$$(1-\beta)V_F(\theta) - C - \delta \geq V_F(\theta) - \beta V_F(\theta_R) - C - \delta - \gamma(\theta_R)a \qquad (5)$$

Equation (5) signifies that the expected payoff of a franchisee who declares actual sales must be higher than the payoff when behaving opportunistically. By adopting a cheating strategy, he can certainly retain a higher share of total revenues yet runs the risk of being detected and punished by the franchiser ($\gamma(\theta_R)a$). After rearrangement, $\gamma(\theta_R)$ should be at least equal to:

$$\gamma(\theta_R) \geq \frac{\beta(V_F(\theta) - V_F(\theta_R))}{a} \qquad (6)$$

If the franchiser wishes to deter all franchisees from behaving opportunistically, then he must announce a monitoring policy defined for any θ_R by:

$$\gamma(\theta_R)=\frac{\beta(V_F(\bar{\theta})-V_F(\theta_R))}{a} \quad (7)$$

Since on a market with local demand $\bar{\theta}$ the franchisee has higher incentive to under-declare, it is sufficient to deter cheating by also preventing the other franchisees with lower demand from cheating.

A purely company-owned system

If the chain is seeking to deter a salaried manager with a local demand of θ from cheating, his probability of supervision $\varpi(\theta_E)$ must be defined by:

$$w-\delta \geq w-\delta e-\varpi(\theta_E)b \quad \text{with}^{18} \quad e=\frac{V_F(\theta_E)}{V_F(\theta)}<1 \quad (8)$$

After rearrangement, the intensity of monitoring $\varpi(\theta_E)$ should be at least equal to:

$$\frac{\delta(V_F(\theta)-V_F(\theta_E))}{bV_F(\theta)} \quad (9)$$

If the chain is seeking to deter all salaried managers from shirking, then it must implement a monitoring policy $\varpi(\theta_E)$, characterized for any θ_E by:

$$\varpi(\theta_E)=\frac{\delta(V_F(\bar{\theta})-V_F(\theta_E))}{bV_F(\bar{\theta})} \quad (10)$$

With such monitoring intensity, managers with the highest incentives to shirk (*i.e.* those experiencing demand $\bar{\theta}$) are deterred, as are all those with a level of demand $\theta<\bar{\theta}$.

[18] Exerting an effort $e=\frac{V_F(\theta_E)}{V_F(\theta)}$ leads to a level of sales $eV_F(\theta)=V_F(\theta_E)$ instead of $V_F(\theta)$.

We will now examine whether the policy of monitoring all stores is really optimal. We will first consider the case of a vertically-integrated chain.

3.2 Purely company-owned chain and partial monitoring

In accordance with Besanko and Spulber (1989), we have considered a two-part monitoring program defined by:

$$\begin{array}{lll} & \theta_E > \hat{\theta}, \quad then & \varpi(\theta_E) = 0 \\ if & & \\ & \theta_E \le \hat{\theta}, \quad then & \varpi(\theta_E) = \delta \dfrac{\left(V_F(\hat{\theta}) - V_F(\theta_E)\right)}{bV_F(\hat{\theta})} \end{array} \qquad (11)$$

Complete monitoring would imply setting $\hat{\theta}=\overline{\theta}$, whereas partial monitoring is characterized by $\hat{\theta} < \overline{\theta}$. In the latter case, the chain does not monitor those managers generating sales greater than $V_F(\hat{\theta})$ and concentrates its resources on monitoring units with sales of below $V_F(\hat{\theta})$. Moreover, the intensity of monitoring tends to decrease with the unit's observed sales. Given the monitoring policy defined in (11), managers experiencing a demand $\theta \le \hat{\theta}$ are deterred from shirking and will always choose $\theta_E=\theta$ (*i.e.* an effort e=1). Managers experiencing $\theta > \hat{\theta}$ know that they will never be monitored as long as they are achieving at least $V_F(\hat{\theta})$. They therefore choose effort level $e = \dfrac{V_F(\hat{\theta})}{V_F(\theta)}$, which enables reaching sales amount $V_F(\hat{\theta})$[19]. With partial monitoring, the chain can be considered as offering a rent to managers with $\theta > \hat{\theta}$. The performance distribution of stores thus tends to be

[19] We can easily prove that a salaried manager experiencing a demand $\theta > \hat{\theta}$ can never expect a higher payoff by exerting an effort below $e = \dfrac{V_F(\hat{\theta})}{V_F(\theta)}$, since: $w - \delta \dfrac{V_F(\theta_E)}{V_F(\theta)} - \left(\dfrac{\delta\left(V_F(\hat{\theta}) - V_F(\theta_E)\right)}{bV_F(\hat{\theta})} \right) b < w - \delta \dfrac{V_F\left(\hat{\theta}\right)}{V_F(\theta)}$, where the expression on the left-hand side represents his expected benefit from cheating in choosing a level of effort of $\theta_E < \hat{\theta}$ and the expression on the right-hand side the expected benefit if he chooses $\theta_E = \hat{\theta}$.

upward bounded, with a concentration of stores displaying the same sales ($V_F(\hat{\theta})$). Given the monitoring policy, the following proposition demonstrates that the partial monitoring strategy is indeed optimal.

Proposition 1: *Within a purely company-owned chain, it is never optimal to monitor all units.*

Proof: We must show that it is optimal to set $\hat{\theta} < \overline{\theta}$. Optimal monitoring policy is determined by maximizing the chain's profit with respect to the monitoring threshold $\hat{\theta}$:

$$G = \int_{\underline{\theta}}^{\hat{\theta}} \left(V_F(\theta) - w - C - \varpi(\theta)K\right) f(\theta)d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \left(V_F(\hat{\theta}) - w - C\right) f(\theta)d\theta \tag{12}$$

The derivative with respect to $\hat{\theta}$ is given by:

$$\frac{\partial G}{\partial \hat{\theta}} = \left[V_F(\hat{\theta}) - w - C - \varpi(\hat{\theta})K - V_F(\hat{\theta}) + w + C\right] f(\hat{\theta}) + V'_F(\hat{\theta}) \int_{\hat{\theta}}^{\overline{\theta}} f(\theta)d\theta + \frac{\partial \varpi}{\partial \hat{\theta}} K \int_{\underline{\theta}}^{\hat{\theta}} f(\theta)d\theta$$

After rearrangement:

$$\frac{\partial G}{\partial \hat{\theta}} = -\int_{\underline{\theta}}^{\hat{\theta}} \frac{\delta V'_F(\hat{\theta}) V_F(\theta) K}{b\left(V_F(\hat{\theta})\right)^2} f(\theta)d\theta + V'_F(\hat{\theta}) \int_{\hat{\theta}}^{\overline{\theta}} f(\theta)d\theta \tag{13}$$

The first expression has a negative sign and corresponds to a *monitoring cost effect*: a higher $\hat{\theta}$ increases the cost of supervising managers since a greater number of units are being monitored. The second expression has a positive sign and represents an *incentive effect*: the positive impact from more intense monitoring on managers' efforts and thus on sales and profits as well. Monitoring all units entails setting $\hat{\theta} = \overline{\theta}$. Such a policy however is not optimal since:

$$\left.\frac{\partial G}{\partial \hat{\theta}}\right|_{\hat{\theta}=\overline{\theta}} = -\int_{\underline{\theta}}^{\overline{\theta}} \frac{\delta V'_F(\overline{\theta}) V_F(\theta) K}{b\left(V_F(\overline{\theta})\right)^2} f(\theta)d\theta < 0 \tag{14}$$

From (14), we conclude that the chain can raise profits by relaxing supervision on the most efficient units: from a level of $\hat{\theta} = \overline{\theta}$, reducing $\hat{\theta}$ by a slight amount enables the chain to significantly lower its monitoring costs (first-order effect), without considerably weakening the incentive for managers to behave honestly (second-order effect).

Moreover, we find that a *laissez-faire* policy ($\hat{\theta} = \underline{\theta}$) is not optimal:

$$\left.\frac{\partial G}{\partial \hat{\theta}}\right|_{\hat{\theta}=\underline{\theta}} = V_F'(\underline{\theta}) > 0 \qquad (15)$$

In the neighborhood of $\underline{\theta}$, by increasing $\hat{\theta}$, the chain can stimulate efforts from all its managers (first-order effect), with the expected benefits largely compensating for the additional monitoring cost. ■

Proposition 1 can be explained as follows. Partial deterrence allows for a significant reduction in monitoring costs. Even if the revenues generated by the chain are less than what would be expected in stores experiencing high demand, this shortfall is more than counterbalanced by the savings on monitoring costs. It should be noted that the "informational rent" captured by managers $\left(\delta - \delta\frac{V_F(\hat{\theta})}{V_F(\theta)}\right)$ is an increasing function of monitoring costs and a decreasing function of penalty severity (as $\hat{\theta}$ decreases in *K* and increases in *b*). The proportion of monitored units should thus be close to 100% as monitoring costs decline towards zero or when punishments become extremely severe. Yet by raising the penalties, two opposite effects are introduced on monitoring intensity at the individual level. Those units subject to a negative demand shock are now less intensively monitored[20], whereas those submitted to a positive shock are more likely to be monitored (even if the probability is

[20] Since manager behavior is only influenced by the expected penalty, a higher penalty enables reducing the probability of audit.

infinitesimal in presence of high demand). Even under these extreme conditions (severe punishments, monitoring costs close to zero) however, it is still efficient for the chain to renounce auditing the most profitable units.

3.3 Purely franchised chain and partial monitoring

Let's now consider the monitoring program for a purely franchised system as follows:

$$\begin{array}{lll} \theta_R > \hat{\hat{\theta}}, & then & \gamma(\theta_R) = 0 \\ if & & \\ \theta_R \leq \hat{\hat{\theta}}, & then & \gamma(\theta_R) = \beta \dfrac{\left(V_F(\hat{\hat{\theta}}) - V_F(\theta_R)\right)}{a} \end{array} \qquad (16)$$

Full monitoring entails setting $\hat{\hat{\theta}} = \overline{\theta}$, whereas when $\hat{\hat{\theta}} < \overline{\theta}$, the chain only monitors a subset of units and "offers" rent to those franchisees experiencing high demand. From a formal standpoint, all franchisees experiencing demand $\theta > \hat{\hat{\theta}}$ can under-declare their revenues ($V(\hat{\hat{\theta}})$ instead of $V(\theta)$) and pay $\beta V_F(\hat{\hat{\theta}})$ instead of $\beta V_F(\theta)$. As long as the royalties paid remain greater than or equal to $\beta V_F(\hat{\hat{\theta}})$, they are not monitored. It is therefore profitable for them to declare revenues equal to $V(\hat{\hat{\theta}})$[21]. Conversely, all franchisees exposed to demand $\theta \leq \hat{\hat{\theta}}$ are monitored and have a strong incentive to declare their actual revenues.

Proposition 2: *In a purely franchised chain, it is never optimal to monitor all stores.*

---

[21] For franchisees of the type $\theta > \hat{\hat{\theta}}$, the benefit from declaring revenues $V_F(\theta_R)$ below $V_F(\hat{\hat{\theta}})$ is given by:

$$V_F(\theta) - \beta V_F(\theta_R) - C - \delta - \left(\frac{\beta\left(V_F(\hat{\hat{\theta}}) - V_F(\theta_R)\right)}{a} \right) a = V_F(\theta) - \beta V_F(\hat{\hat{\theta}}) - C - \delta$$

Hence, they are perfectly indifferent when declaring between $V_F(\hat{\hat{\theta}})$ and lower revenues.

Proof in the Appendix.

The rationale for partial monitoring in a purely franchised system is the same as in a company-owned system. The existence of monitoring costs incites the chain to refrain from auditing the most profitable franchised units. A franchisee experiencing demand $\theta > \hat{\hat{\theta}}$ will thus benefit from an "informational rent" equal to $\beta\left(V_F(\theta) - V_F(\hat{\hat{\theta}})\right)$[22]. Offering a rent to the higher-earning franchisees represents a cost-saving method for the franchiser to deter franchisees from engaging in more harmful cheating (a more detrimental action would be to declare the lowest level of sales $V_F(\underline{\theta})$). We have noticed that partial monitoring incites the franchiser to audit only honest franchisees, yet this is still necessary if the chain wishes to prevent "high sales" franchisees from passing themselves off as "low sales" franchisees.

Up until now, we have shown that complete monitoring is never optimal in pure chains. Would monitoring policy exhibit the same pattern in the presence of dual distribution?

3.4 Dual distribution and efficiency

The aim of this article is to determine whether a purely franchised chain (respectively, a vertically-integrated chain) can benefit from introducing company-owned units (respectively, franchised units). In order to address this issue, we first need to establish which managers will volunteer to become franchisees or salaried staff should a pure chain switch to dual distribution. The underlying idea is that a contract (either a franchise or labor contract) will be signed only if the two parties derive mutual benefit from doing so (a salaried manager cannot be coerced into becoming a franchisee or a franchisee into becoming a salaried employee). The chain holds two levers (royalties and manager wages) to increase or decrease the relative attractiveness of franchisee and salaried statuses.

Let's now consider a purely company-owned chain. By setting reasonable royalty rates, the chain can incite some of its salaried managers to apply for franchising. These applicants however are more likely to be present on local markets experiencing high demand, given that that franchisee gains increase with expected sales. Similarly, in a purely franchised chain, the franchiser can pressure some of his franchisees into a salaried position by raising royalty rates: the volunteers will be those encountering the least favorable demand conditions[23].

Lemma: *For reasonable values of w and β, there exists a level of demand* $\tilde{\theta}$ *defined by:*

$$w - \delta \frac{V_F(\hat{\theta})}{V_F(\tilde{\theta})} = (1-\beta)V_F(\tilde{\theta}) - C - \delta \qquad (17)$$

such that any manager experiencing a local demand above $\tilde{\theta}$ *will apply for a franchise, while those with a demand below* $\tilde{\theta}$ *will opt for being salaried.*

$\tilde{\theta}$ is the level of demand that leaves a manager indifferent between being a franchisee or salaried. The expression on the right-hand side of equation (17) corresponds to the profit of a franchisee experiencing a demand $\tilde{\theta}$ and who behaves honestly, knowing he will be monitored, because he will be declaring the lowest revenues from among all franchised units (with the other franchised units experiencing a demand between $\tilde{\theta}$ and $\bar{\theta}$). The expression on the left-hand side corresponds to the gains realized by a salaried manager experiencing demand $\tilde{\theta}$ who is aware of his status of not being monitored (since he outperforms the rest of the company-owned units). This manager can therefore furnish an effort $e = \frac{V_F(\hat{\theta})}{V_F(\tilde{\theta})} < 1$, where $V_F(\hat{\theta})$ is the threshold sales level below which the chain initiates monitoring. In the following

[22] The greater the franchisee's demand, the higher the rent he may receive.

proposition, we will prove that a vertically-integrated chain can always increase profits by introducing franchised units.

Proposition 3: *If a chain is completely company-owned, then franchising some units would be optimal.*

Proof: Consider that the pure chain sets a royalty rate β such that all salaried managers experiencing demand $\theta>\tilde{\theta}$ wish to become franchisees and all managers with $\theta\leq\tilde{\theta}$ prefer to remain company-owned, with $\tilde{\theta}$ defined by: $w-\delta\frac{V_F(\hat{\theta})}{V_F(\tilde{\theta})}=(1-\beta)V_F(\tilde{\theta})-C-\delta$.

If the chain decides to sign a franchise contract with all applicants, then the chain's profit will be as follows:

$$G=\int_{\underline{\theta}}^{\hat{\theta}}(V_F(\theta)-w-C-\varpi(\theta)K)f(\theta)d\theta+\int_{\hat{\theta}}^{\tilde{\theta}}\left(V_F\left(\hat{\theta}\right)-w-C\right)f(\theta)d\theta \\ +\int_{\tilde{\theta}}^{\hat{\hat{\theta}}}(\beta V_F(\theta)-\gamma(\theta)Q)f(\theta)d\theta+\int_{\hat{\hat{\theta}}}^{\bar{\theta}}\beta V_F(\hat{\hat{\theta}})f(\theta)d\theta \tag{18}$$

The first two expressions correspond to the profits earned on the company-owned units (see (12)) and the last two to the earnings on franchised units (see (22)). If an interior solution $\tilde{\theta}$ were to exist that maximizes the chain's profit, then it would be given by:

$$\frac{\partial G}{\partial\tilde{\theta}}=\left(V_F\left(\tilde{\theta}\right)-w-C\right)f(\tilde{\theta})-\left[\beta V_F(\tilde{\theta})f(\tilde{\theta})-\gamma(\tilde{\theta})Qf(\tilde{\theta})\right]=0 \tag{19}$$

The optimal proportion of company-owned units $\tilde{\theta}$ is obtained by means of comparing the profit from a company-owned unit run by an unsupervised manager and the royalties generated from a monitored franchisee minus the supervision cost. The optimal level of franchise mix is such that the gain from an additional franchised unit is equal to the profit from an additional company-owned unit.

[23] The franchiser can alternatively increase the wage offered to salaried managers.

At $\tilde{\theta} = \bar{\theta}$ (characterizing a purely company-owned chain), we have:

$$\left.\frac{\partial G}{\partial \tilde{\theta}}\right|_{\tilde{\theta} = \bar{\theta}} = \left(V_F(\hat{\theta}) - C - w\right) f(\bar{\theta}) - \beta V_F(\bar{\theta}) f(\bar{\theta}) \qquad (20)$$

since $\gamma(\bar{\theta}) = 0$ (*i.e.* if an additional marginal franchisee were to be introduced, there would be no need to monitor since local demand is perfectly known by the chain). This derivative corresponds to the income difference between a company-owned store and a franchised store, with both being run under the most favorable demand conditions.

Since $w - \delta \frac{V_F(\hat{\theta})}{V_F(\bar{\theta})} = (1-\beta) V_F(\bar{\theta}) - C - \delta$, then:

$$\left.\frac{\partial G}{\partial \tilde{\theta}}\right|_{\tilde{\theta} = \bar{\theta}} = -\left(V_F(\bar{\theta}) - V_F(\hat{\theta})\right)\left(1 - \frac{\delta}{V_F(\bar{\theta})}\right) \qquad (21)$$

From Assumption 1, we know that $\delta < V_F(\bar{\theta})$.

Hence, $\left.\frac{\partial G}{\partial \tilde{\theta}}\right|_{\tilde{\theta} = \bar{\theta}} < 0$.

A purely company-owned chain is never optimal and the chain can increase profits by introducing some franchised units. ■

By allowing a few franchisees to integrate into the chain, the franchiser reduces asymmetric information on the demand; he is most certainly aware that the recruited franchisees will evolve in a very favorable market environment with a demand towards the upper range of the random distribution. As a result, it will be useless for the franchiser to perform intense monitoring for disciplinary purposes. Moreover, these franchisees will exert pressure upon the salaried managers and enable the chain owner to reduce the rent it would be required to pay them. For these reasons, the chain will save on monitoring costs by transforming some company-owned units into franchised units. This organizational switch

however implies setting reasonable royalty rates (if royalties are too high, applicants will prefer the status of salaried manager).

Proposition 4: *If a chain is completely franchised, then converting some franchised units into company-owned units would be optimal.*

Proof in the Appendix.

The role of company units is to limit the opportunism of franchisees and prevent them from declaring excessively low sales. By converting some franchises into company units, the franchiser can therefore save on monitoring costs. From the two previous propositions, we can directly infer the central proposition of our model: a dual system is more efficient than a purely franchised system or a purely company-owned system.

Proposition 5: *When a chain cannot perfectly observe local demands and manager behavior, then it is always optimal to hire both franchised and salaried managers.*

The parameter ($\tilde{\theta}$) can be viewed as the ideal level of franchising mix. According to Propositions 3 and 4, this target company ownership level always lies in the interval $\left]\underline{\theta},\overline{\theta}\right[$ and depends on: monitoring policy, wage rates, and required royalty rates. By modifying monitoring intensity within the company-owned stores $\hat{\theta}$ or the level of royalty rates β, the chain can indeed increase or reduce the extent of company ownership ($\tilde{\theta}$). Therefore, the franchiser can manipulate at least two interdependent strategic variables to target a proportion of the company-owned units: monitoring intensity $\hat{\theta}$, and the level of royalty rates β.

Corollary: *Decreasing monitoring intensity in company-owned units and increasing royalty rates (β) are complementary instruments in the aim of increasing the percentage of company-owned units.*

This result yields some of the theoretical foundations that enable proving a positive relationship between royalty rates and the proportion of company-owned units, as exhibited in several previous empirical studies (Lafontaine, 1992; Lafontaine and Shaw, 1999; Pénard, Raynaud and Saussier, 2003). Such a postulate suggests that these two instruments may be interlinked to best govern chains, thus confirming the findings of Bradach (1997) regarding restaurant chains.

3.5 Discussion

Our theoretical framework has yielded insight into the predominance and efficiency of dual distribution in franchising. Our model lends support to the rationale that dual distribution heavily depends upon the monitoring policy, which is a somewhat original thesis with respect to the literature available on the topic. Our results rely to an extent on the assumption that the franchiser cannot identify the actual state of local demand without incurring costs and that franchisees have an incentive to under-declare their actual revenues (and save on royalties). Even if franchisees might be tempted to under-declare their actual sales, this would not necessarily imply that they all systematically will. Their decision will depend on the risk of being detected and punished. Some may decide to go ahead and cheat (if they estimate that the gain is higher than the expected cost) while others may behave honestly. Even if all franchisees are tempted to cheat, revenue chiseling at equilibrium will be minimal as long as monitoring costs remain low.

Several implications may be drawn from our model. First, a chain will rely more

intensively on franchising when it possesses more advanced technology to monitor its franchised units. Conversely, the proportion of company-owned units should decrease with the monitoring costs of salaried managers and increase with the severity of punishments that the chain is able to inflict upon opportunistic employees.

Second, if manager effort is not critical to increasing sales, then concern over the shirking hazard becomes less important, and we should observe more vertical integration in order to concentrate on mitigating the under-declaration hazard. This premise would seem to be consistent with empirical results. For instance, Lafontaine (1992) found that the proportion of franchised units tends to rise when the franchisee's effort counts; conversely, the extent of vertical integration is more prevalent for chains where franchisee efforts are less crucial. Similarly, if the prospect of sales under-declaration is smaller than the shirking problem or is more easily controlled, we should then observe greater reliance on franchising.

Lastly, monitoring intensity may also depend on contractual provisions. It has been argued in the literature, for instance, that special provisions such as tying may lower monitoring costs or improve monitoring technology (Klein and Saft, 1985; Michael, 2000). This feature can be illustrated in the case where the chain provides one or several inputs to franchisees. If the production function at the store is of a fixed proportion, then the chain is able to infer the actual level of sales from the quantity of inputs sold to a particular outlet (*i.e.* the chain can infer the true level of output and avoid "excessive" discrepancies between output and declared sales). Since the discretional behavior of franchisees is limited, we would expect the monitoring intensity of franchised units to be lower with greater reliance on franchising in a chain with a tying provision.

4. CONCLUSION

Most of the existing theoretical literature on franchising considers dual distribution as a transitory phenomenon, while empirical results show both its prevalence and stability. Our paper has suggested that dual distribution is an efficient and stable organizational choice when both local demand and manager behavior are costly to observe. Dual distribution enables chains to save on outlet monitoring costs and to mitigate the moral hazard.

Our findings on the efficiency of dual distribution can however be understood from a more general perspective: an organization experiencing the kind of contractual hazard we have focused on herein is always more efficient by mixing the incentive schemes made available to its members or agents, *i.e.* by combining fixed and revenue-based remuneration. In other words, the policy of generalizing variable remuneration within an organization might be as sub-optimal as relying exclusively on fixed remuneration. If we were to assimilate variable remuneration with external procurement and fixed remuneration with internal procurement, our results could also be used as another rationale for the prevalence of what is sometimes called "tapered integration" in vertical relationships, *i.e.* the simultaneous use of external input suppliers and in-house suppliers.

Lastly, our propositions may be related to the discussion on optimal monitoring policy within organizations in general and retail chains in particular. As stated in the introduction, most papers on franchising typically assume that sales or output observation is costless, which incites chains to link franchisee compensation with sales. This constitutes a form of output monitoring. In company-owned units, chain managers are provided fixed wages and directly supervised by the chain. Chains are thus relying on input (or behavior-based) monitoring within their own units. The optimality of dual distribution implies that it is efficient to exercise both input and output monitoring at the chain level even if individual units will not

be monitored at both the input and output levels. Input and output monitoring thereby coexist in a plural chain even if each of these devices is "specialized" for a particular type of individual unit. Those units whose manager is paid with a fixed compensation are input-monitored, whereas those offering variable compensation are output-monitored: this set-up is consistent with the way chains are organized and with previous results by Khalil and Lawarée (1995) on the choice between input and output monitoring[24].

[24] In their paper, they found that when the principal (the chain in this instance) is the residual claimant (who owns the outlet), he generally prefers input monitoring, but output monitoring would be preferred when the agent (here, the franchisee) is the residual claimant (owns the outlet). See also Bontems and Bourgeon, 2000).

APPENDIX

Proof of Proposition 2: The optimal intensity of monitoring $\hat{\hat{\theta}}$ is given by maximizing the franchiser's profit with respect to managerial threshold monitoring $\hat{\hat{\theta}}$:

$$Max_{\left\{\hat{\hat{\theta}}\right\}} G = \int_{\underline{\theta}}^{\hat{\hat{\theta}}} (\beta V_F(\theta) - \gamma(\theta)Q) f(\theta) d\theta + \int_{\hat{\hat{\theta}}}^{\overline{\theta}} \beta V_F(\hat{\hat{\theta}}) f(\theta) d\theta \qquad (22)$$

The derivative of the profit with respect to $\hat{\hat{\theta}}$ is given by:

$$\frac{\partial G}{\partial \hat{\hat{\theta}}} = -\gamma(\hat{\hat{\theta}}) Q f(\hat{\hat{\theta}}) - \beta \frac{V'_F(\hat{\hat{\theta}})}{a} Q \int_{\underline{\theta}}^{\hat{\hat{\theta}}} f(\theta) d\theta + \beta V'_F(\hat{\hat{\theta}}) \int_{\hat{\hat{\theta}}}^{\overline{\theta}} f(\theta) d\theta \qquad (23)$$

The first two expressions represent the negative *monitoring cost effect* if the chain decides to tighten monitoring of its franchisees (to increase $\hat{\hat{\theta}}$). The third expression corresponds to the *incentive effect* (more intensive monitoring induces franchisees to declare their actual sales). Complete monitoring consists of setting $\hat{\hat{\theta}} = \overline{\theta}$. Such a policy however is not optimal since the franchiser is able to increase profit by relaxing monitoring efforts on the most efficient units. Marginal profit is found to decrease with $\hat{\hat{\theta}}$ at $\hat{\hat{\theta}} = \overline{\theta}$:

$$\left.\frac{\partial G}{\partial \hat{\hat{\theta}}}\right|_{\hat{\hat{\theta}} = \overline{\theta}} = -\beta \frac{V'_F(\overline{\theta})}{a} Q < 0 \qquad (24)$$

We can also notice that a *laissez-faire* policy $\hat{\hat{\theta}} = \underline{\theta}$ is sub-optimal:

$$\left.\frac{\partial G}{\partial \hat{\hat{\theta}}}\right|_{\hat{\hat{\theta}} = \underline{\theta}} = \gamma(\underline{\theta}) Q f(\underline{\theta}) + \beta V'_F(\underline{\theta}) > 0 \qquad (25)$$

Proof of Proposition 4: At $\widetilde{\theta} = \underline{\theta}$ (characterizing a purely franchised chain) and according to equation (23), we have:

$$\left.\frac{\partial G}{\partial \widetilde{\theta}}\right|_{\widetilde{\theta} = \underline{\theta}} = (V_F(\underline{\theta}) - C - w) f(\underline{\theta}) - \left(\beta V_F(\underline{\theta}) f(\underline{\theta}) - \gamma(\underline{\theta}) Q f(\underline{\theta})\right) \tag{26}$$

As shown previously, this derivative corresponds to the difference of incomes between a company-owned store and a franchised store, both run under the worst market conditions (*i.e.* lowest level of demand).

Since $w - \delta \frac{V_F(\underline{\theta})}{V_F(\underline{\theta})} = (1-\beta) V_F(\underline{\theta}) - C - \delta$, we then obtain:

$$\left.\frac{\partial G}{\partial \widetilde{\theta}}\right|_{\widetilde{\theta} = \underline{\theta}} = \gamma(\underline{\theta}) Q f(\underline{\theta}) > 0 \tag{27}$$

REFERENCES

Azevedo P.F., and V.L. Dos Santos Silva., 2001, "Contractual Mix Analysis in The Brazilian Franchising", Unpublished Manuscript.

Bai, C.E, and Z. Tao, 2000, "Contract Mix in Franchising", *Journal of Economics and Management Strategy*, 9(1), 85-113.

Baron, D.P., and D. Besanko, 1984, "Regulation, Asymmetric Information and Auditing", *RAND Journal of Economics*, 15, 447-470.

Becker, G.S, 1968, "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 76(2), 169-217.

Bercovitz, J.E., 2001, "The Organizational Choice Decision in Business-Format Franchising: An Empirical Test", working paper, Fuqua School of Business, Duke University.

Besanko, D., and D. Spulber, 1989, "Antitrust Enforcement under Asymmetric Information", *Economic Journal*, 99, 408-425.

Bhattacharyya, S., and F. Lafontaine, 1995, "Double-Sided Moral Hazard and the Nature of Share Contracts", *RAND Journal of Economics*, 26(4), 761-781.

Bontemps, P., J-M. Bourgeon, 2000, "Creating Countervailing Incentives through the Choice of Instruments", *Journal of Public Economics*, 76, 181-202.

Bradach, J.L., 1997, "Using the Plural Form in the Management of Restaurant Chains", *Administrative Science Quarterly*, 42, 276-303.

Brickley, J., and F.H. Dark, 1987, "The Choice of Organizational Form: The Case of Franchising", *Journal of Financial Economics*, 18, 401-420.

Brickley, J.A., 1999, "Incentive Conflicts and Contractual Restraints: Evidence from Franchising", *Journal of Law & Economics*, XLII, 745-774.

Brickley, J.A., 2002, "Royalty Rates and Upfront Fees in Share Contracts: Evidence from

Franchising", *Journal of Law, Economics & Organization*, 18(2), 511-535.

Bull, J., and J. Watson, 2004, "Evidence Disclosure and Verifiability", *Journal of Economic Theory*, 118, 1-31.

Caves, R.E., and W.F. Murphy, 1976, "Franchising: Firms, Markets and Intangible Assets", *Southern Economic Journal*, 42(4), 572-586.

Combs, J.G., S.C. Michael, G.J. Castrogiovanni, 2004, "Franchising: A Review and Avenues to Greater Theoretical Diversity", *Journal of Management*, 30(6), 907-931.

Gal-Or, E., 1995, "Maintaining Quality Standard in Franchised Chains", *Management Science*, 41(11), 1774-1792.

Gallini, N.T., and N. Lutz, 1992, "Dual Distribution and Royalty Fees in Franchising", *Journal of Law, Economics & Organization*, VIII, 471-501.

Holmström, B., and P. Milgrom, 1991, "Multitask Principal Agent Analysis: Incentive Contracts, Asset Ownership and Job Design", *Journal of Law, Economics & Organization*, 7, 24-52.

Khalil, F., 1997, "Auditing without Commitment"', *RAND Journal of Economics*, 28(4), 629-640.

Khalil, F., J. Lawarée, 1995, "Input versus Output Monitoring: Who is the Residual Claimant?", *Journal of Economic Theory*, 66, 139-157.

Klein, B., 1995, "The Economics of Franchise Contracts", *Journal of Corporate Finance*, 2, 9-37.

Klein, B., L.F. Saft, 1985, "The Law and Economics of Franchise Tying Contracts", *Journal of Law & Economics*, XXVIII, 345-361.

Krueger, A.B., 1991, "Ownership, Agency and Wages: An Examination of Franchising in the Fast Food Industry, *Quarterly Journal of Economics*, 106(1), 75-101.

Lafontaine, F., 1992, "Agency Theory and Franchising: Some Empirical Results", *RAND*

Journal of Economics, 23, 263-283.

Lafontaine, F., 1993, "Contractual Arrangement as Signaling Devices: Evidence from Franchising", *Journal of Law, Economics & Organization*, 9, 256-289.

Lafontaine, F., and E. Raynaud, 2002, "Residual Claims and Ongoing Rent as Incentive Mechanisms in Franchise Contracts: Complements or Substitutes?", in E. Brousseau and J-M. Glachant, eds., *The Economics of Contracts: Theories and Applications*, Cambridge University Press, 315-336.

Lafontaine, F., K.L. Shaw, 1999, "The Dynamics of Franchise: Evidence from Panel Data", *Journal of Political Economy*, 107, 1041-1080.

Lafontaine, F., and K.L. Shaw, 2005, "Targeting Managerial Control: Evidence from Franchising", *RAND Journal of Economics*, 36, 131-150.

Lafontaine, F., and M. Slade, 2001, "Incentive Contracting and the Franchise Decision", in K. Chatterjee and W. Samuelson (eds*.), Game Theory and Business Applications*, Kluwer Academic Press, 133-188.

Lafontaine, F., and M. Slade, 1996, "Retail Contracting and Costly Monitoring: Theory and Evidence", *European Economic Review*, 40, 923-932.

Lewin-Solomon, S.B., 1999, "Innovation and Authority in Franchise Systems: An Exploration of the Plural Form", working paper, Dept. of Applied Economics, Cambridge University.

Martin, R.E., 1988, "Franchising and Risk Management", *American Economic Review*, 78, 954-968.

Mathewson, F., R. Winter, 1985, "The Economics of Franchise Contracts", *Journal of Law & Economics*, XXVIII(3), 503-526.

Minkler, A., 1990, "An Empirical Analysis of a Firm Decision to Franchise", *Economic Letters*, 34, 77-82.

Michael, S.C., 2000, "The Extent, Motivation and Effect of Tying in Franchise Contracts", *Managerial and Decision Economics*, 21, 191-201.

Mookherjee, D., and I.P.L. Png, 1989, "Optimal Auditing, Insurance, and Redistribution", *Quarterly Journal of Economic*, 104, 339-415.

Norton, S., 1988, "An Empirical Look at Franchising as an Organizational Form", *Journal of Business*, 61, 197-217.

Oxenfeldt, A.R., and A.O. Kelly, 1969, "Will Successful Franchise Systems Ultimately Become Wholly-Owned Chains?" *Journal of Retailing*, 44, 69-83.

Pénard T., E. Raynaud and S. Saussier, 2003, "Dual Distribution and royalty rates in Franchised Chains: An Empirical Analysis Using French Data", *Journal of Marketing Channels*, 10, 5-32.

Rubin, P., 1978, "The Theory of the Firm and the Structure of Franchise Contract", *Journal of Law & Economics*, XXI(1), 223-233.

Scott, F.A., 1995, "Franchising vs. Company Ownership as a Decision Variable of the Firm", *Review of Industrial Organization*, 10, 69-81.

Shapiro, C., J.E. Stiglitz, 1984, "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, 74(3), 433-444.

Sorenson O., and J.B. Sorensen, 2001, "Finding the Right Mix: Organizational Learning, Plural Forms, and Franchise Performance", *Strategic Management Journal*, 22 (6-7), 713-724.