An Adaptation Theory Of Franchising*

Giorgio Zanarone†

September 12, 2006

Abstract

Several authors have analysed franchise contracts as static incentive mechanisms designed to motivate the parties to exert high managerial effort via an appropriate specification of the royalty rate. In this paper, I take a complementary approach and I study how the long-term nature of franchise agreements is exploited to induce opportunistic franchisees to implement local decisions (observable to the parties but non-contractible) that are efficiently adapted to a changing environment. In a world of incomplete contracts and limited freedom to terminate them at will, adaptation can be achieved by allocating formal decision powers to the franchisors, who can use them, ex post, to choose which decisions the franchisees are supposed to implement and which remedies to apply against non-compliance. If the parties are sufficiently concerned about the future of their relationship, under an appropriate allocation of decision rights, franchising replicates vertical integration where that best performs (controlling observable local decisions) and, by making franchisees residual claimants of the downstream profits, it succeeds where vertical integration fails (motivating franchisees to exert high unobservable effort).

1 Introduction

Like integration, franchising has widely been studied in the organizational economics literature, despite the absence of a commonly accepted definition of what makes it especially attractive as a way to govern vertical relationships. In an archetypical franchise contract, a firm (the franchisor) grants another firm (the franchisee) the right to sell goods or services under her brand name. While the franchisee's advantage from signing such a contract is well understood (she saves on the costs of signalling the quality of her products to the final customers[1]), the

*I thank Benito Arruñada, Robert Gibbons and Giuseppe Zanarone for valuable advise, Petar Balachev, Laura Ferrari Bravo, Walter Garcia Fontes, Manuel Gonzalez-Diaz, Paolo Mariti, Antonio Nicita, Markus Reitzig, Stephane Saussier, Michael Sykuta and Ernesto Villanueva for helpful comments.

†Department of Economics and Business, Universitat Pompeu Fabra and MIT, Sloan School of Management (visiting). E-mail: giorgio.zanarone@upf.edu.

[1] Rubin (1978) mentions two other potential benefits a franchisee obtains from the franchise contract. First, he may receive ongoing training from the franchisor and, second, he may gain

franchisor's gains are more disputable. In fact, the franchisor bears the risk that another, independent firm, with partially diverging interests, will jeopardize her reputational investments by reducing the local quality standards to save on operating costs. In other words, franchise contracts are plagued by externalities, whose magnitude may vary depending on how different the parties' objectives are but will generally be well above zero. In this perspective, a theory of franchising should essentially explain why and when franchisors find it attractive to put their reputation in different hands, rather than nurturing it behind the arguably safer bars of a vertically integrated distribution chain. Such theoretical question can be decomposed into three, intimately related sub-questions: what organizational advantages does franchising have over integration? On what instruments, if any, can franchisors rely to mitigate the externality-driven costs of losing direct control on their reputation? What are the characteristics of the economic and institutional environment that limit and shape the use of such instruments? This paper attempts to answer the three questions above and, while so doing, it aims to develop a unified theory of the choice between franchising and vertical integration and of the optimal design of franchise contracts.

In an article published in 1995 on the Journal of Corporate Finance, Benjamin Klein proposes a simple tradeoff between vertical integration and franchising. According to Klein's argument, vertical integration enables manufacturers to control the observable managerial decisions implemented at the retail outlets, at the cost of muting the incentives of local managers to exert high amounts of unobservable effort in the completion of productive tasks. Conversely, franchising provides retailers with "naturally" powerful effort incentives, since they are rewarded with the residual profits of the operations they manage, but this occurs at the cost of increasing the retailers' incentives to adopt opportunistic managerial decisions, since, by owning the downstream assets, they control all those decisions to whose implementation the assets are essential. There's no explicit mention, however, in Klein's paper, of the possibility of conveying decision control to the franchisors by assigning them decision and enforcement powers rather than via vertical integration. The omission may[2] reflect the author's implicit convinction that such possibility is not a plausible one, an argument that is more explicitly sustained by Lafontaine and Shaw (2003), when they write: "Franchisors rely on various contractual restraints to curb free-riding, such as requiring some minimum advertising and service levels or developing strict product mix guidelines. But such restraints are costly to enforce and still may not prevent all the behaviors that franchisors worry about. Those with more valuable brands then may need to operate more outlets directly so they can exercise more direct managerial control over downstream operations". The

easier access to the credit market as a result of the franchisor's willingness to guarantee for him in front of creditors.

The relevance of the first benefit, as Rubin himself acknowledges, depends on the comparative advantage of the franchisor with respect to the market in supplying ongoing training. As regards the second benefit, it is closely related to the quality signalling advantage I mention in the text, since both derive from the franchisor's superior reputation.

[2] ...or may not. To this respect, see Klein and Murphy (1988).

existing empirical evidence, however, indicates that both the direct control of retail outlets through vertical integration[3] and the indirect control mediated by restrictive franchise contracts[4] are used to contrast the franchise externalities .

In Section 3 of this paper, I appropriate Klein's intuition and I argue that integration, by conveying to the manufacturers real control over the downstream operations and residual claimancy of both upstream and downstream profits, represents a direct, *static* solution to the problem of vertical governance, which comes at the cost of muting the effort incentives faced by retail managers. In Section 4, I attempt to go beyond this tradeoff between vertical integration and outsourcing, and I argue that the assignment of decision powers to the franchisors, which is one of the most recurrent features of franchise contracts, serves, like integration, the purpose of neutralizing the franchise externalities, while preserving, at the same time, the effort incentives associated with outsourcing. While the decision control conveyed by integration is direct, however, the one conveyed by franchise contracts is indirect and mediated by the institutional system that regulates the enforceability of contractual obligations. When such system does not work effectively, I show that franchise contracts can still facilitate the adoption of efficient managerial decisions by bundling the decision rights of franchisors with the power to terminate non-complying franchisees. In this perspective, franchising can be regarded as an essentially relational governance structure, whose effectiveness in combining the advantages of integration and outsourcing depends on the extent to which the parties care about the continuation of their business relation.

The rest of the paper is organized as follows. In Section 2, I describe the interaction between a manufacturer and the manager of a retail outlet, who trade with each other in an environment characterized by rapidly changing market conditions, imperfect third-party enforcement and the existence of legal constraints on the parties' ability to terminate their relationship at will. In Sections 3 and 4, I explore the economic tradeoffs between integration and non-integration in a static environment. In Section 5, I move to a relational environment and I analyse the tradeoffs between unrestricted and restricted franchise contracts. Under restricted franchising, the *legal* power to select and adapt the decisions to be implemented by the franchisees and the right to terminate the contract in case of non-compliance is assigned to the franchisor. In the last paragraph of Section 5, I discuss the comparative statics of the model, some testable implications and the way they can be reconciled with the existing empirical evidence. Finally, in Section 6, I consider the possibility that, under restricted franchising, a franchisor endowed with superior bargaining power uses his fomal decision rights to enforce an exploitative relational contract, in which the franchisee is required to implement decisions that are privately optimal for the franchisor but inefficient from the point of view of the vertical chain as a whole, and I briefly discuss the testable predictions that this enrichment of the theory generates and the strategies that could be followed to empirically disen-

[3] See, for instance, Arruñada and Vázquez (2003), Brickley, Dark and Weisbach (1991), Lafontaine (1992) and Lafontaine and Shaw (2005).

[4] See, among others, Arruñada, Garicano and Vázquez (2001, 2005) and Brickley (1999).

tangle an eficiency hypothesis from an exploitation hypothesis on the adoption of restricted franchising as a form of vertical governance.

2 The economic and institutional environment

In this model, there are two risk-neutral parties (a manufacturer, M, and a distributor, D, who markets M's products to the final consumers) and two assets (the upstream asset, which is used in the production process, and the downstream asset, which is used in the distribution process). M exogenously owns the upstream asset, whereas the downstream asset can be either owned by M or D. I assume that asset ownership conveys asset-inseparable payoffs, which I denote, respectively, by π_U and π_D[5].

Two types of actions are relevant in the management of the downstream asset: the adoption of *observable decisions* (for instance, fixing the size of sales and after-sales staff, the product mix, the retail prices, the location of outlets) and the allocation of *unobservable effort* to productive tasks (for example, spending time and attention in promoting the product and describing its characteristics to customers or supplying after-sales services such as repairing)[6]. In this paper, I focus on a simple environment in which there's only one dimension of the downstream operations on which decisions must be taken and a unique task to the completion of which effort and attention must be directed[7].

The effect of a given decision and effort level on the asset-specific payoffs in each period depends on which state of the world, s, is drawn from the set, S, of all possible states[8]. The payoff functions attached the to upstream and downstream assets are denoted by $\pi_U(d, e, s)$ and $\pi_D(d, e, s)$, where π_U and π_D are separable in d and e, $\pi_D(e = 0, s) = 0$, $\frac{\partial \pi_D}{\partial e}(s) > 0$ and $\frac{\partial^2 \pi_D}{\partial e^2}(s) < 0$ for every s. Effort has a private, state-independent cost $C(e)$, where

$$\begin{cases} C(e) = C'(e) = 0 & \text{for } 0 \leq e \leq \underline{e} \\ C'(e) > 0, C''(e) > 0 & \text{for } e > \underline{e}^{9} \end{cases}$$

[5] The subscripts U and D stand for "upstream" and "downstream", respectively.

[6] The fact that, in most franchise chains, certain tasks that are delegated to the franchisee can hardly be monitored due to the distance of franchisees' outlets from the headquarters, has received considerable attention in both the theoretical and empirical literature. See, for instance, Rubin (1978), Norton (1988) and Lafontaine (1992).

[7] A useful way of interpreting the effort variable in this model is, following Holmstrom and Milgrom (1991), in terms of time and attention devoted to enhance the value of the downstream asset, where such value is non-contractible but transferable via ownership of the asset. A similar interpretation of asset ownership as a way to provide effort incentives is adopted in Baker and Hubbard (2004).

[8] The important role that the need to adapt business decisions to the economic environment plays in driving the choice of organizational arrangements has been emphasized in the theoretical contributions of Simon (1951), Williamson (1979, 1985, 1991), Grossman and Hart (1987), Baker, Gibbons and Murphy (2004, 2006) and Gibbons (2005), as well as in the empirical works of Masten and Crocker (1986), Crocker and Reynolds (1993), Lerner and Merges (1999), Arruñada, Garicano and Vázquez (2001), Elfenbein and Lerner (2003), Baker and Hubbard (2004) and Forbes and Lederman (2005).

. Hence, the total surplus generated by the transaction between M and D can be written as

$$TS = \pi_U(d, e, s) + \pi_D(d, e, s) - C(e) \tag{1}$$

. In what follows, I make a set of assumptions on how the decision and the level of effort are chosen, ex post, by the parties. First, decisions, though observable, are ex ante and ex post non-contractible[10]. Second, the *real power* to implement a state-contingent decision $d(s)$ belongs to the owner of the downstream asset. This implies that, for instance, if M owns the downstream asset, he will be able to directly implement, ex post, any decision he considers appropriate, without needing D's cooperation (which would be crucial if D, instead, owned the asset). Third, the power to exert effort inalienably belongs to D[11]. Fourth, M and D can assign by contract the *formal right* to select a decision ex post. In that case, the party who is assigned the decision right has the power to burden the other party with a contractual obligation to comply, ex post, with her preferred decision. Fifth, the only enforceable sanction the party who is assigned a decision right can invoke against the other party's non-compliance is the termination of the contract. Sixth, termination "at will" is not allowed, so the only way of terminating a contract is in response to the breach of an explicit obligation[12].

3 The efficient benchmark

In a first best world, in which effort is observable and contracts are complete and perfectly enforceable, M and D would agree, in any state s, on the decision and the effort level that solve

$$\underset{d,e}{Max} \{\pi_U(d, e, s) + \pi_D(d, e, s) - C(e)\} \tag{2}$$

[10] The ex ante non-contractibility assumption is at the core of the property rights theories of the firm. See, for instance, Grossman and Hart (1986), Hart and Moore (1990), Halonen (2002), Baker, Gibbons and Murphy (2002) and Matouschek (2004). The ex post non-contractibility assumption is made in Baker, Gibbons and Murphy (2004, 2006), Hart and Moore (2006) and Hart and Holmstrom (2002).

[11] At this stage, D could either be the owner of the downstream asset or an employee, so no assumptions are made (yet) about D's payoff function. What is assumed is that, no matter what the organizational structure and the underlying incentives are, M cannot personally exert the effort that is necessary to the functioning of downstream operations and, therefore, he must delegate work to D, who will, then, select the amount of effort she is willing to exert according to her incentives.

[12] In many legal systems, "at will" termination can be enforced only if the termination decision is notified one or more years in advance and, perhaps more interestingly, if fair compensatory payments are payed to the terminated party, where fairness is defined by a judge, perhaps along the guidelines provided by the law. In practice, these limitations often make "at will" termination very costly and, therefore, they discourage its use as an enforcement device. On this, see Epstein (1975), Brickley, Dark and Weisbach (1991) and Paz-Ares (1999).

. Since the payoff functions are separable in d and e, (2) can be decomposed into the subproblems

$$\underset{d}{Max}\left\{\pi_U(d,s)+\pi_D(d,s)\right\} \tag{2'}$$

$$\underset{e}{Max}\left\{\pi_U(e,s)+\pi_D(e,s)-C(e)\right\} \tag{2"}$$

, where the necessary and sufficient first order condition for (2") is

$$\frac{\partial \pi_U\left(e^{FB},s\right)}{\partial e}+\frac{\partial \pi_D\left(e^{FB},s\right)}{\partial e}=\frac{\partial C\left(e^{FB}\right)}{\partial e} \tag{3}$$

. Denote the solution of (2') by $d^{FB}(s)$ and the solution of (2") by $e^{FB}(s)$. Ex ante, the first best expected total surplus will be

$$TS^{FB}=E_s\left[\pi_U\left(d^{FB}(s),e^{FB}(s),s\right)+\pi_D\left(d^{FB}(s),e^{FB}(s),s\right)-C\left(e^{FB}(s)\right)\right] \tag{4}$$

4 Static governance

Under static governance, M and D interact only once. The timing of the game is as follows: in the first stage, M and D choose a governance form, that is, they negotiate an assignment of the ownership of the downstream asset. In the second stage, a state s, drawn from the finite set S, is realized. In the third stage, the owner of the downstream asset observes the realization of s and, given that, he chooses which decision to implement. At the same time, D observes s and, given that, she decides how much effort to exert. Finally, in the fourth stage, M and D's payoffs are realized.

Given that, in the absence of the threat of termination, formal decision rights are unenforceable, in a static environment, the owner of the downstream asset is free to implement his own preferred decision. Therefore, there are two possible static governance structures:

1) non-integration, in which D owns the downstream asset and, *therefore*, decides;

2) integration, in which M owns the downstream asset and, *therefore*, decides.

4.1 Non-integration

Under non-integration, D chooses, ex post, both d and e. In each state, D's problem can be decomposed into

$$\underset{d}{Max}\left\{\pi_D(d,s)\right\} \tag{5}$$

$$\underset{e}{Max}\left\{\pi_D(e,s)-C(e)\right\} \tag{6}$$

, where the necessary and sufficient first order condition for (6) is

$$\frac{\partial \pi_D\left(e^{D},s\right)}{\partial e}=\frac{\partial C\left(e^{D}\right)}{\partial e} \tag{7}$$

. Denote the solution of (5) by $d^D(s)$ and the solution of (6) by $e^D(s)$. Ex ante, the expected total surplus under static non-integration will be

$$TS^{NI} = E_s\left[\pi_U\left(d^D(s), e^D(s), s\right) + \pi_D\left(d^D(s), e^D(s), s\right) - C\left(e^D(s)\right)\right] \quad (8)$$

. Notice that, as long as there are some states in which $d^D(s) \neq d^{FB}(s)$ or $e^D(s) \neq e^{FB}(s)$ (that is, states in which M and D's interests are not perfectly aligned and a moral hazard problem exists), the total surplus under outsourcing is smaller than the first best one ($TS^{NI} < TS^{FB}$).

4.2 Integration

Under integration, ex post, M chooses d and D chooses e. In each state, M's and D's problems are, respectively

$$\underset{d}{Max}\left\{\pi_U(d,s) + \pi_D(d,s)\right\} \quad (9)$$

$$\underset{e}{Max}\left\{W - C(e)\right\} \quad (10)$$

, where W is the compensation that M pays to D when D is the manager of M's vertically integrated distribution outlet. The fact that D's compensation is independent of her choice of effort depends on the assumptions that π_U and π_D are asset-inseparable (i.e., no sharing contracts are feasible) and that there are no contractible variables correlated with e[13], which can be used as proxies for D's level of effort. The solution of (9) is $d^{FB}(s)$ as in the first best case. The necessary and sufficient first order condition for (10) is

$$\frac{\partial C(e)}{\partial e} = 0 \quad (11)$$

, so the amount of effort exerted by D is $\underline{e}$, where $\underline{e}$ is the maximum level of effort such that $C'(\underline{e}) = 0$. Ex ante, the expected total surplus under integration will be

$$TS^I = E_s\left[\pi_U\left(d^{FB}(s), \underline{e}, s\right) + \pi_D\left(d^{FB}(s), \underline{e}, s\right)\right] \quad (12)$$

Notice that, as long as there are some states in which $\underline{e} < e^{FB}(s)$, $TS^I < TS^{FB}$.

4.3 Static second best: integration Vs non-integration

The static model envisions a simple tradeoff between integration and non-integration. On one hand, integration internalizes the externalities that typically induce D to choose the decision $d(s)$ without regard for its effect on M's

[13] This case is similar to the one envisioned in Holmstrom and Milgrom (1991), where the agent's effort in enhancing the value of the principal's asset is described as unobservable to both the principal and any third party (that is, fully non-contractible).

profits[14]. On the other hand, integration does not allow to replicate the effort incentives D faces under non-integration for being residual claimant of the profits attached to the downstream asset, and this causes D to exert minimal effort in every state. This is consistent with Klein's informal argument, according to which the cost of integration relative to franchising is its reduced ability to provide incentives for those components of the distributors' performance that are partially or totally unobservable. Given that no contractible proxies for D's effort exist in this model and, therefore, an organizational structure that induces D to exert $e^{FB}(s)$ in each state is not feasible, the optimal choice of governance structure in the static environment can be summarized through the following

Proposition 1 *Denote the decision-related and effort-related components of the total surplus under integration and non-integration by*

$$TS_d^I = E_s\left[\pi_U\left(d^{FB}(s), s\right) + \pi_D\left(d^{FB}(s), s\right)\right] \tag{13}$$

$$TS_d^{NI} = E_s\left[\pi_U\left(d^{D}(s), s\right) + \pi_D\left(d^{D}(s), s\right)\right] \tag{14}$$

$$TS_e^I = E_s\left[\pi_U\left(\underline{e}, s\right) + \pi_D\left(\underline{e}, s\right)\right] \tag{15}$$

$$TS_e^{NI} = E_s\left[\pi_U\left(e^{D}(s), s\right) + \pi_D\left(e^{D}(s), s\right) - C\left(e^{D}(s)\right)\right] \tag{16}$$

. Suppose $TS_e^{NI} \leq TS_e^I$. Then, integration will be the optimal static governance structure for the relationship between M and D. Suppose, conversely, that $TS_e^{NI} > TS_e^I$. Then , integration will be optimal if, and only if $TS_d^I - TS_d^{NI} > TS_e^{NI} - TS_e^I$.

Proof. *By inspection of (13) through (16).* ■

Proposition 1 states that, when high effort is unimportant or even detrimental to the good management of the downstream asset[15], or when the unobservable component of the distributor's performance is less valuable than the observable component, integration will result as a more efficient static governance structure than non-integration.

5 Relational governance: a case for franchising?

Can we find an organizational structure that combines the advantages of integration (guaranteeing the implementation of $d^{FB}(s)$ in each state) with those

[14] Brickley, Dark and Weisbach (1991), Lafontaine and Show (2005) and Forbes and Lederman (2005) provide evidence consistent with the idea that vertical integration enables manufacturers to better control distributors' decisions and, this way, internalize vertical externalities. Interestingly, both papers are less convincing and neat in describing the costs of intergation.

[15] The fact that $e^D(s) \geq \underline{e}$ for any s does not imply that the level of effort $e^D(s)$ is more efficient than $\underline{e}$. If e is interpreted as the effort exerted by D in maintaining the value of the downstream asset, for example, it is easy to understand why $e^D(s)$ could be greater than $e^{FB}(s)$, that is, why, under outsourcing, D could be induced to exert effort *in excess* with respect to the first best. In fact, D might underuse her asset (that is, direct too much effort to value-preserving activities) for failing to take into account the effect of a given level of asset utilization on M's profits.

of non-integration (guaranteeing that D exerts effort $e^D(s) > \underline{e}$ in each state)? Such an organizational structure, if feasible, will generate the second best (expected) total surplus

$$TS^{SB} = E_s\left[\pi_U\left(d^{FB}(s), e^D(s), s\right) + \pi_D\left(d^{FB}(s), e^D(s), s\right) - C\left(e^D(s)\right)\right] \tag{17}$$

, thereby improving on both integration and non-integration. We have seen that, under static governance, the tradeoff between efficient decisions and low effort is unescapable and, therefore, TS^{SB} cannot be achieved. The reason is that, under integration, decisions are efficiently controlled by M but, given that effort is non-contractible, D cannot be motivated to work beyond the minimal level $\underline{e}$, while, under static outsourcing, the residual claimancy of the downstream profits induces D to raise, in each state, her effort up to $e^D(s)$, but the ownership of the downstream asset endows D with decision control, thus allowing her to implement, in each state, $d^D(s)$ rather than $d^{FB}(s)$.

In this section, I study the choice of governance structure when M and D try to achieve the second best through a long term, relational contract. It is worth remarking that no relational contract can lead to TS^{SB} when M and D are integrated. In fact, static integration already guarantees that the efficient decision is implemented, which implies that a relational contract would only improve on the static outcome by inducing D to exert a level of effort higher than $\underline{e}$. However, a relational contract is not court-enforceable and, therefore, can be implemented only if the parties are able to detect breach. Since effort is unobservable, M cannot distinguish, say, between a case in which D has exerted minimal effort but the market conditions have been "good" and a case in which D has exerted the promised level of effort but the market conditions have been "bad". This means that M cannot base any punishment strategy on D's failure to exert the promised amount of effort and, therefore, a relational contract requiring D to exert greater-than-minimal effort will never be self-enforcing. Non-integration, however, can, potentially, form the basis for a second best relational contract. On one hand, non-integration provides D with (static) incentives to exert, in each state, the high level of effort $e^D(s)$. On the other hand, once the state of the world is revealed, both parties obseve the decision implemented by D and, therefore, a deviation from the efficient decision schedule $d^{FB}(s)$ can be detected and punished.

5.1 Unrestricted Vs. restricted franchising

I define a relational governance structure as an allocation of the ownership of the downstream asset *and* the legal decision right. Hence, there are two possible relational governance structures: *unrestricted franchising*, under which D owns the asset and has the decision right, and *restricted franchising*[16], under which

[16] The term "franchising" is used, here, in a non-legal sense and, therefore, doesn't imply the payment of a royalty and a franchise fee from D to M. All my definition of franchising erquires is a) the existence of a long-term relationship between M and D and b) the presence of externalities that misalign M and D's interests. I believe this definition captures some

D owns the asset and M has the decision right. In this paragraph, I use the results in Levin (2003) and Baker, Gibbons and Murphy (2006) to study the trigger strategy equilibria of the repeated game between M and D under both relational governance structures.

Assuming M and D repeat their transaction indefinitely, the timing of each period's game is the following. At stage 0, M and D jointly allocate the ownership of the downstream asset and the legal decision right. At the beginning of stage 1, M pays D a fixed sum, $w \gtreqless 0$. At the end of stage 1, a state of the world, $s \in S$, is realized. At stage 2, M publicly declares the decision he wants D to implement. At the beginning of stage 3, D implements either M's decision or her own preferred decision, $d^D(s)$ and exerts effort $e^D(s)$. At the end of stage 3, if D has implemented M's decision, M decides whether to pay D a bonus, $B(d(s)) \geq 0$. If D has implemented the opportunistic decision $d^D(s)$, M pays no bonus. Finally, at stage 4, M and D earn their respective payoffs, π_U and π_D, as a function of the choices made at stage 3. If either party reneges on the terms of the relational contract at stage 3, M and D revert to static governance from the following period and thereafter.

The difference between unrestricted and restricted franchising is that, once a breach of the relational contract is detected, restricted franchising enables M to terminate D, while unrestricted franchising does not[17]. Hence, the two relational governance structures generate different streams of punishment payoffs and, consequently, different conditions for the self-enforceability of the second best relational contract. The validity of this point rests on two assumptions. First, M and D are unable to renegotiate the ownership of the downstream asset after the relational contract is broken and, therefore, the per period total surplus in the punishment phase is TS^{NI} under unrestricted franchising and $TS^T = p_M + p_D$ under restricted franchising, where p_M and p_D are M and D's per period excpected payoffs after termination[18]. The rationale for this

important common features of long-term manufacturer-distributor relationships, which extend from franchise contracts in a strictly legal sense (Mc Donald's, Starbuck's Coffee, etc.) to similar forms of vertical arrangements, where the payment of royalties and franchise fees is not always observed (car dealerships). The analysis could (and perhaps will) be extended to include the use of royalties as a form of bilateral monetary incentive. My choice to abstract from this aspect here depends on the fact that the incentive properties of royalties have been the main focus of the economic literature on franchising, whereas my purpose here is to enlighten the role of control rights as an instrument, perhaps complementary to the payment of royalties, to achieve adaptive decision-making without having to resort to vertical integration.

[17] Recall that, in the institutional environment modeled in this article, M cannot terminate D "at will". Therefore, termination can only be enforced if M is explicitly assigned the right to terminate D for failure to comply with an obligation or upon occurrence of a contractually specified circumstance.

The decision/termination rights studied in this model belong to the first category (termination for non-compliance). In fact, they result from the combination of contractual provisions assigning M the right to specify state-contingent obligations for D and a general legal rule that allows a party in a contract to terminate the other party for failure to comply with an obligation that forms part of the contract.

[18] A similar assumption is made in Halonen (2002).

assumption, which is non-standard in the literature[19], is that a relational contract is based on the existence of mutual trust, which is lost if some party does not honor the terms of the agreement. Second, M is better off terminating D, after the latter proves to be untrustworthy, rather than continuing to deal with her, while D prefers to continue to operate as a franchisee for M rather than being terminated. Formally, this means that $p_M > E_s\left[\pi_U\left(d^D(s), e^D(s), s\right)\right]$ and $p_D < E_s\left[\pi_D\left(d^D(s), e^D(s), s\right) - C\left(e^D(s)\right)\right]$. The foundation for this assumption lies in Klein's time-honored argument, according to which franchise contracts are generally used to discipline asymmetric relationships. According to this approach, the competitiveness of the market for franchises and the fact that the franchisees make specific investments in the course of their relation with a franchisor assure that the franchisees' outside options are relatively more unattractive than the franchisors' ones, and this, in turn, makes the threat of termination from the franchisors an effective enforcement instrument[20].

Define

$$E_M^{SB} = E_s\left[\pi_U\left(d^{FB}(s), e^D(s), s\right) - B\left(d^{FB}(s)\right)\right] - w \quad (18)$$

$$E_D^{SB} = E_s\left[\pi_D\left(d^{FB}(s), e^D(s), s\right) + B\left(d^{FB}(s)\right) - C\left(e^D(s)\right)\right] + w \quad (19)$$

$$E_M^{NI} = E_s\left[\pi_U\left(d^D(s), e^D(s), s\right)\right] \quad (20)$$

$$E_D^{NI} = E_s\left[\pi_D\left(d^D(s), e^D(s), s\right) - C\left(e^D(s)\right)\right] \quad (21)$$

. Given the assumptions above, under unrestricted franchising, M and D will be willing to honor the relational contract if, respectively,

$$\frac{1+r}{r}E_M^{SB} \geq \frac{1+r}{r}E_M^{NI} \quad (22)$$

$$\frac{1+r}{r}E_D^{SB} \geq \frac{1+r}{r}E_D^{NI} \quad (23)$$

$$-B\left(d^{FB}(s)\right) + \frac{1}{r}E_M^{SB} \geq \frac{1}{r}E_M^{NI} \quad (24)$$

$$\pi_D\left(d^{FB}(s), s\right) + B\left(d^{FB}(s)\right) + \frac{1}{r}E_D^{SB} \geq \pi_D\left(d^D(s), s\right) + \frac{1}{r}E_D^{NI} \quad (25)$$

, $\forall\ s \in S$[21]. Summing up the individual rationality constraints (22) and (23) yields the necessary condition

$$TS^{SB} > TS^{NI} \quad (26)$$

[19] Several recent models of relational contracting assume that, after breach is detected, the parties revert to the optimal static governance structure (see Baker, Gibbons and Murphy (2001, 2002, 2004, 2006) and Levin (2003)). This is probably why these models do not emphasize the role of termination in facilitating the self-enforceability of relational contracts.

[20] See Klein (1980), Klein and Murphy (1988) and Klein (1995). Similar arguments are made in Haldfield (1990), Brickley, Dark and Weisbach (1991), Paz-Ares (1999) and Arruñada, Garicano and Vázquez (2001, 2005).

[21] Theoretically, a third self-enforceability condition should be added, namely that D doesn't have an incentive to refuse the bonus offered by M. However, since the bonus is paid after D implements $d^{FB}(s)$, refusing the bonus after foregoing the opportunity to implement the opportunistic decision $d^D(s)$ rather than $d^{FB}(s)$ is a dominated strategy for D. Therefore, condition (9) is sufficient for D not to have an incentive to refuse the bonus.

, which is satisfied by definition (the second best surplus is greater than the non-integration surplus).

As noted by Levin (2003), if (24) and (25) hold for every $s \in S$, they must also hold for the state $\overline{s}$ for which M and D have the strongest temptation to renege on the terms of the relational contract. Therefore, the $2S$ constraints implied by (24) and (25) can be reduced to the 2 constraints

$$-B\left(d^{FB}\left(\overline{s}\right)\right)+\frac{1}{r}E_{M}^{SB} \geq \frac{1}{r}E_{M}^{NI} \quad (24')$$

$$\pi_{D}\left(d^{FB}\left(\overline{s}\right),\overline{s}\right)+B\left(d^{FB}\left(\overline{s}\right)\right)+\frac{1}{r}E_{D}^{SB} \geq \pi_{D}\left(d^{D}\left(\overline{s}\right),\overline{s}\right)+\frac{1}{r}E_{D}^{NI} \quad (25')$$

. Summing up constraints (24') and (25') and solving for the interest rate r yields the unique necessary condition

$$r \leq \frac{TS^{SB}-TS^{NI}}{\pi_{D}\left(d^{D}\left(\overline{s}\right),\overline{s}\right)-\pi_{D}\left(d^{FB}\left(\overline{s}\right),\overline{s}\right)} \quad (27)$$

. The same analysis can be applied to show that the unique necessary condition for the relational contract to be self-enforcing under restricted franchising (i.e., when M has the power to terminate D for non-compliance) is

$$r \leq \frac{TS^{SB}-TS^{T}}{\pi_{D}\left(d^{D}\left(\overline{s}\right),\overline{s}\right)-\pi_{D}\left(d^{FB}\left(\overline{s}\right),\overline{s}\right)} \quad (28)$$

. Theorem 3 in Levin (2003) assures that, if M and D have no liquidity constraints (that is, if they have sufficient liquidity to make any individually rational side-payment to each other) (27) and (28) are also sufficient conditions for the relational contract to be self-enforcing. In other words, if (27) and (28) hold, then there exist values of w and $B\left(d^{FB}\left(\overline{s}\right),\overline{s}\right)$ such that the other self-enforcement conditions hold as well. An inspection of (27) and (28) proves the following

Proposition 2 *Define the most efficient relational governance structure as the one that makes the second-best relational contract self-enforcing for the largest range of interest rates. Then, if total surplus is higher under termination than under non-integration ($TS^{T} > TS^{NI}$), unrestricted franchising will be more efficient than restricted franchising. Conversely, if total surplus is higher under non-integration than under termination ($TS^{NI} > TS^{T}$), restricted franchising will be more efficient than unrestricted franchising.*

The main testable prediction suggested by Proposition 2 is that, when D's assets are relationship-specific, (that is, when TS^{T} is small relative to TS^{NI}), restricted franchising is likely to be optimal.

5.2 Comparative statics

Inequalities (27) and (28) have the same denominator, $\Delta\pi_{D}\left(\overline{s}\right)=\pi_{D}\left(d^{D}\left(\overline{s}\right),\overline{s}\right)-\pi_{D}\left(d^{FB}\left(\overline{s}\right),\overline{s}\right)$, which represents D's maximum reneging temptation, that is,

D's gain from implementing the opportunistic decision $d^D(\overline{s})$ in the state in which such gain is highest[22]. This suggests that, as $\Delta\pi_D(\overline{s})$ diminishes, the second best relational contract is more likely to be self-enforcing under both restricted and unrestricting franchising. On the other hand, when $\Delta\pi_D(\overline{s})$ increases, it becomes more and more likely that the relational contract will be self-enforcing *only* under the governance structure that makes the self-enforcement condition loosest. However, if D's reneging temptation grows too large, the second best relational contract will not be self-enforcing under either governance structure. In this case, M and D will try to implement a "third best" relational contract, that is, a contract that generates a smaller maximum reneging temptation and a smaller total surplus. This process will eventually stop when the best relational contract M and D can implement yields a total surplus not greater than the ex ante optimal static surplus. In that case, at stage 0, M and D will choose to either integrate, engage in static non-integration or stay separate, depending on which of these solutions yields the greatest total surplus[23].

The above intuition can be formalized as follows. Solving for D's maximum reneging temptation, we can rewrite conditions (27) and (28), respectively, as

$$\Delta\pi_D(\overline{s}) \leq \frac{TS^{SB} - TS^{NI}}{r} \tag{27'}$$

$$\Delta\pi_D(\overline{s}) \leq \frac{TS^{SB} - TS^{T}}{r} \tag{28'}$$

. Denote the right end sides of (27') and (28'), respectively, by $\overline{\Delta\pi_{DSB}}^{UF}$ and $\overline{\Delta\pi_{DSB}}^{RF}$. Suppose termination constitutes the strongest punishment ($TS^{NI} > TS^{T}$). This implies that $\overline{\Delta\pi_{DSB}}^{RF} > \overline{\Delta\pi_{DSB}}^{UF}$, that is, the self-enforcement condition is looser under restricted franchising. Finally, define $TS^{Sep} = P_M + P_D$ as the total surplus M and D achieve if, at stage 0, they decide not to initiate a business relationship. We can, now, prove the following

Proposition 3 *Assume $TS^{NI} > TS^{T}$. Then, i) if $\Delta\pi_D(\overline{s}) < \overline{\Delta\pi_{DSB}}^{UF}$, the second best relational contract will be self-enforcing under both restricted and unrestricted franchising and, therefore, unrestricted franchising will be chosen[24]; ii)*

[22] On this, see, also, Baker, Gibbons and Murphy (2004, 2006).

[23] At stage 0, M and D trust each other and, therefore, they can reach an agreement on the optimal static governance structure. Whether such an agreement will result in integration, non-integration or separation depends on the tradeoff between adaptive decisions (assured by integration) and high effort (motivated by non-integration), as well as on the parties' outside options at stage 0, P_M and P_D.

[24] The underlying assumption, here, is that choosing restricted franchising (i.e., assigning a decision/termination right to M) involves an unmodeled contracting cost, which can be thought of as small enough to be irrelevant when the adoption of restricted franchising makes self-enforcement possible, but large enough to discourage the adoption of restricted franchising when self-enforcement is also guaranteed by unrestricted franchising.

The cost of restricted franchising might be the simple "ink cost" of specifying M's decision/termination right in a contract, or it might result, more interestingly, from the parties' concern that an asymmetric contract could be looked upon suspiciously by third party enforcers, such as antitrust authorities.

if $\overline{\Delta\pi}_{DSB}^{UF} < \Delta\pi_D(\overline{s}) < \overline{\Delta\pi}_{DSB}^{RF}$*, the second best relational contract will be self-enforcing only under restricted franchising, which will be, therefore, the chosen governance structure; iii) if* $\Delta\pi_D(\overline{s}) > \overline{\Delta\pi}_{DSB}^{RF}$*, the second best relational contract will not be self-enforcing under either governance structure. In that case, M and D will choose the relational governance structure that achieves the highest level of total surplus,* TS^{TB}*, such that* $TS^{TB} \geq Max\left\{TS^{NI}, TS^{I}, TS^{Sep}\right\}$*. If such a governance structure does not exist, M and D will adopt the best static governance structure, that is, the one that generates total surplus* $TS^{TB} = Max\left\{TS^{NI}, TS^{I}, TS^{Sep}\right\}$*.*

5.3 Testable predictions

In a special case, Proposition 3 can generate neat testable predictions on M and D's choice of governance structure. Suppose that, when $\Delta\pi_D(\overline{s}) > \overline{\Delta\pi}_{DSB}^{RF}$ and M and D resort to the third best best relational contract, the decrease in D's maximum reneging temptation, denoted by $\Delta\pi_D(\overline{s}) - \Delta\pi_D\left(\overline{\overline{s}}\right)$, is smaller than the decrease in the total surplus, $TS^{SB} - TS^{TB}$, for any $\overline{\overline{s}} \neq \overline{s}$. This would be the case, for instance, if the states of the world occurred with relatively similar probabilities, the adoption of efficient decisions in each state were highly valuable to M and the variation in D's reneging temptation across states were small. In such an environment, assuming that a third best relational contract improving on static governance is feasible, it will be that $\overline{\Delta\pi}_{DTB}^{UF} < \Delta\pi_D\left(\overline{\overline{s}}\right) < \overline{\Delta\pi}_{DTB}^{RF}$ and, therefore, restricted franchising will still be the efficient relational governance structure when $\Delta\pi_D(\overline{s}) > \overline{\Delta\pi}_{DSB}^{RF}$. When no third best relational contract is feasible, it will be efficient for M and D to agree on the best static governance, which does not involve the assignment of formal decision power to M. Suppose, now, that, given the best feasible relational contract, D's maximum reneging temptation is given by the random variable

$$\Delta\pi_D = X + u \tag{29}$$

, where $E[u] = 0$. Then, the variable X is an unbiased, linear proxy for $\Delta\pi_D$. Let

$$RF = \begin{cases} = 1 & \textit{if restricted franchising is chosen} \\ = 0 & \textit{otherwise} \end{cases}$$

. Then, according to Proposition 3,

$$\begin{aligned} E[RF] &= prob(RF = 1) = prob\left(\overline{\Delta\pi}_{DTB}^{UF} < \Delta\pi_D < \overline{\Delta\pi}_{DTB}^{RF}\right) = \\ &= prob\left(\overline{\Delta\pi}_{DTB}^{UF} - X < u < \overline{\Delta\pi}_{DTB}^{RF} - X\right) = \\ &= prob\left(u < \overline{\Delta\pi}_{DTB}^{RF} - X\right) - prob\left(u < \overline{\Delta\pi}_{DTB}^{UF} - X\right) = \\ &= F\left(\overline{\Delta\pi}_{DTB}^{RF} - X\right) - F\left(\overline{\Delta\pi}_{DTB}^{UF} - X\right) \end{aligned} \tag{30}$$

, where F is the cumulative distribution function for u. We can now determine how the probability to observe restricted franchising varies with X:

$$\frac{\partial prob\left(RF=1\right)}{\partial X}=f\left(\overline{\Delta\pi}_{DTB}^{UF}-X\right)-f\left(\overline{\Delta\pi}_{DTB}^{RF}-X\right) \tag{31}$$

. This implies that

$$\frac{\partial prob\left(RF=1\right)}{\partial X}\begin{cases} >0 & if\ f\left(\overline{\Delta\pi}_{DTB}^{UF}-X\right)>f\left(\overline{\Delta\pi}_{DTB}^{RF}-X\right) \\ <0 & otherwise \end{cases} \tag{32}$$

. Under reasonable non-monotonicity assumptions about the form of $f\left(u\right)$, (32) leads to predict that, given any unbiased proxy X for D's reneging temptation, the probability to observe restricted franchising increases in X for low enough values of X, while it decreases in X for high enough values of X.

The prediction in (32) is partially consistent with the evidence presented by Arruñada, Garicano and Vázquez (2001) and Zanarone (2005), who study cross sections of contracts between car manufacturers and franchised dealers, respectively, from Spain and Italy, and find that, as the dealers' incentives to implement opportunistic decisions[25] (proxied by the size of the dealership networks) increase, the likelihood of a set of formal decision rights being assigned to the manufacturers increases as well. However, the analysis in these papers is limited to those manufacturer-dealer relationships that remain franchised, while no attention is paid to the effect of further increases in the franchise externalities on the choice between franchising and vertical integration. Another body of evidence that could be rivisited in the light of (32) is offered by Brickley, Dark and Weisbach (1991), who study the effect of laws limiting the freedom of franchisors to terminate franchisees for non-performance on the choice between franchising and vertical integration. The authors find out that, in those North American states where franchisors' termination rights are limited by the law, the likelihood of local outlets to be directly managed by the manufacturer increases. The same prediction can be derived from the model developed in this paper, assuming that vertical integration is the statically optimal governnace structure (that is, $TS^{I}>Max\left\{TS^{NI},TS^{Sep}\right\}$), that achieving first best decisions in every state is very important and that, overall, terminating the relationship generates a loss in expected total surplus (that is, $TS^{NI}>TS^{T}$)[26]. In this setting, limiting franchisors' termination rights is equivalent to issuing a legal provision, which mandates a vis-a-vis choice between unrestricted franchising

[25] Examples of dealers' opportunistic decisions are overpricing the automobiles sold to final consumers or saving on the costs of providing pre-sales services and promotion.

[26] The first assumption is implicitly made by Brickley, Dark and Weisbach (1991), since they consider an efficiently performing franchise contract and vertical integration as the only two feasible governance structures. In other words, they do not even mention the possibility that outsourcing with negative externalities might be still preferable to integration, *although*, to motivate the paper, they have to assume that integration yields lower surplus than efficient franchising (that is, they recognize that integration involves a tradeoff between the benefits of increased control and some unspecified cost).

and integration. Therefore, the probability to observe franchising (restricted or unrestricted) is

$$prob\left(UF=1\right)=prob\left(\Delta\pi_{D}<\overline{\Delta\pi}_{DTB}^{UF}\right) \tag{33}$$

when termination laws are present and

$$prob\left(UF=1 \text{ or } RF=1\right)=prob\left(\Delta\pi_{D}<\overline{\Delta\pi}_{DTB}^{RF}\right) \tag{34}$$

when termination laws are not present. Since $\overline{\Delta\pi}_{DTB}^{RF}>\overline{\Delta\pi}_{DTB}^{UF}$, it follows that $prob\left(\Delta\pi_{D}<\overline{\Delta\pi}_{DTB}^{UF}\right)<prob\left(\Delta\pi_{D}<\overline{\Delta\pi}_{DTB}^{RF}\right)$ and, therefore, the introduction of termination laws decreases the probability that franchising is chosen and increases the probability that integration is chosen.

6 Restricting franchising backfires

6.1 Franchisor's opportunism

So far, I have discussed the existence of self-enforcing relational franchise contracts, in which D implements the first best decision ($d^{FB}(s)$) in every state and exerts the high level of effort $e^{D}(s)$. According to Proposition 1, these contracts are efficient in a second best sense, provided that high levels of D's effort are preferable to low levels. In the previous section, I have identified the conditions that make the second best outcome more easily sustainable when a legal decision/termination right is assigned to to M and I have named such a governance structure "restricted franchising". When efficient equilibria are considered, it is appropriate to affirm that governance structures that facilitate their achievement are optimal and, in this sense, Proposition 2 not only identifies the conditions at which restricted franchising is effective, but also the conditions at which it is efficient.

However, when the distribution of M's products is outsourced to D (and, therefore, the downstream profits accrue to D), there are decisions that M (weakly) prefers to the first best ones, namely those decisions that, given the state s, maximize $\pi_{M}(d,s)$ rather than $\pi_{M}(d,s)+\pi_{D}(d,s)$. If M's preferred decision schedule, $d^{M}(s)$, were implemented in each state, the total surplus would be

$$TS^{M}=E_{s}\left[\pi_{U}\left(d^{M}(s),e^{D}(s),s\right)+\pi_{D}\left(d^{M}(s),e^{D}(s),s\right)-C\left(e^{D}(s)\right)\right]<TS^{SB} \tag{35}$$

. Theoretically, M could induce D to enter a relational contract identical to the one analysed in the previous section, except for the fact that the new contract requires D to implement the decision schedule $d^{M}(s)$ instead of $d^{FB}(s)$. In this case, when $TS^{NI}>TS^{T}$, the choice of restricted franchising facilitates not only the achievement of the second best equilibrium, but also the achievement of an inefficient equilibrium, in which M has his preferred, opportunistic

decision schedule, $d^M(s)$, implemented, even though this causes a reduction of total surplus from TS^{SB} to TS^M. Indeed, this is the objection usually raised against explanations of franchisors' unilateral decision powers as contractual arrangements aimed to increase the surplus of the vertical chain as a whole: who guarantees that, once assigned the decision power, franchisors do not use it to enforce privately beneficial, but inefficient decisions?

To qualify this argument and test its soundness, it is important to notice that, as long as M and D have no liquidity constraints and have complete information (at least in the sense that they trust each other's declarations), there is no reason to expect them to honor a relational contract leading to TS^M. In fact, since $TS^M < TS^{SB}$, M and D can, in each period, make fixed transfers to each other, which they will use to split the greater expected surplus TS^{SB}[27]. It seems, therefore, that a theory of restricted franchising as an instrument to enforce M's privately optimal decisions requires to assume the existence of constraints limiting the parties' ability to make the side payments necessary to enforce the efficient relational contract.

6.2 Franchising under asymmetric liquidity constraints

In this paragraph, I assume that D does not have sufficient liquidity to make side payments to M. Following the analysis in section 4, the necessary conditions for the second best relational contract to be self-enforcing are still (27) and (28), whereas the necessary conditions for the exploitative relational contract to be self-enforcing are

$$r \leq \frac{TS^M - TS^T}{\pi_D\left(d^D, \overline{s}\right) - \pi_D\left(d^M, \overline{s}\right)} \tag{36}$$

$$r \leq \frac{TS^M - TS^{NI}}{\pi_D\left(d^D, \overline{s}\right) - \pi_D\left(d^M, \overline{s}\right)} \tag{37}$$

, respectively, under restricted and unrestricted franchising[28]. Notice that, because of D's liquidity constraint, these conditions are not sufficient. However, let me assume, for the sake of argument, that there exist side payments such that, if (36) and (37) are satisfied, then all the other self-enforcement conditions are also satisfied. In that case, conditions (36) and (37) lead to the same result as Proposition 2 on the *effectiveness* of restricted franchising (i.e., on its ability to facilitate the self-enforcement of a given relational contract). The remarkable difference is that, allowing for the possibility that the parties implement relational contracts generating the surplus TS^M, an *effective* governance structure is not necessarily *efficient* (i.e., it does not necessarily facilitate the self-enforcement of the best feasible relational contract).

[27] This property of relational contracts when the parties have deep pockets is due to Levin (2003).

[28] Recall that, by assumption, $p_D < E_s\left[\pi_D\left(d^D, e^D, s\right) - C\left(e^D(s)\right)\right]$ and $p_M > E_s\left[\pi_U\left(d^D, e^D, s\right)\right]$.

6.3 Implications for testability and further enrichments

An important implication of the above analysis regards the (re)interpretation of Prediction (32). In Section 4.3, I have shown that, under certain conditions, the likelihood to observe restrcited franchising as the governance structure adopted by M and D solely depends on D's maximum temptation to renege on the best feasible relational contract. However, in Section 5.2, I have also shown that, when D faces a liquidity constraint, M might be able to induce D to implement an exploitative relational contract, yielding total surplus TS^M. Clearly, D's temptation to renege on M's preferred feasible relational contract will differ from $\Delta\pi_D$ in (14) in all those states in which $d^M(s) \neq d^{FB}(s)$. However, it seems problematic, if possible at all, to empirically distinguish between the two reneging temptations, since the same externalities that give D incentives to deviate from $d^{FB}(s)$ are likely to give her incentives to deviate from $d^M(s)$[29]. Thus, when D's temptation to renege on both the exploitative and the best feasible relational contract is measured by a unique empirical proxy $\widehat{\Delta\pi_D}$, evidence of the type presented in Arruñada, Garicano and Vázquez (2001) and Zanarone (2005) might both support an efficiency hypothesis and an exploitation hypothesis on the adoption of restricted franchising.

The theory, however, opens the way to potential enrichments, which would help to distinguish between the two hypotheses. So far, I have defined M's payoffs as solely dependent on d, e and s, but one could well imagine that M's expected stream of future payoffs also depends on his ability to attract and keep motivated franchisees into his network and, consequently, on his reputation of reliability and fairness in dealing with them. In this case, if M's exploitation (that is, M's success in having $d^M(s)$ implemented in each state via a relational contract) can be easily detected by or credibly communicated to prospective franchisees and if franchisees, ex ante, have relatively appealing choices other than transacting with M, requiring the implementation of $d^M(s)$ in every state

[29] This point may be better clarified by an example. Economic theory suggests that franchisees have an incentive to free ride on the pre-sales services provided by neighbor franchisees belonging to the same network, such as local promotions and demonstrations (see Telser (1960), Brickley and Dark (1987) and, for a critical reinterpretation, Klein and Murphy (1988) and Klein (1996)). Some authors have identified in the density of franchise networks a major factor that positively affects the temptation of franchisees to free ride (see, for instance, Lafontaine (1992), Arruñada, Garicano and Vázquez (2001) and Arruñada and Vázquez (2004)). Franchisees' local promotional activities and demonstrations benefit the franchisor (they increase clients' fidelity and expose a greater number of potential buyers to the product), whereas they both generate benefits (they increase local sales, but to a lower extent than total sales, since part of the incremental sales is captured by intrabrand competitors) and costs (diverting resources from other forms of investments, keeping costly product inventories to offe demonstrations and free trials) for the franchisees. Therefore, the franchisor, if he could, would require franchisees to invest an excsessive amount of resources in advertising and demonstrations, whereas franchisees would invest an excessively low amount of resources. The optimal promotional budget (i.e., the one that maximizes joint profits) would be lower than the one preferred by the franchisor and higher than the one preferred by the franchisees. Clearly, an increase in the network density would affect the temptation of franchisees to deviate from both the first best promotional budget and the franchisor's privately optimal budget, because it would reduce the marginal benefits of an increase in promotional expenditures by increasing the marginal portion of incremental sales captured by intrabrand competitors.

might be too costly for M.

The above argument could be modeled by imagining a "meta-relational contract", whose parties are M, D and a third party, A, which can be thought of as the population of prospective franchisees or, simply, "the market". A is able to observe the decisions commanded by M and the decisions actually implemented by D in each state, as well as M's behavior after a given decision is implemented by D. M and D's strategies are the same as before, while A's strategy can be described as follows. If M terminates D after D's failure to comply with $d^{FB}(s)$, for any $s \in S$, then A remains inert; conversely, if M terminates D after D implements $d^{FB}(s)$, for any $s \in S$, then A initiates an informational campaign against M, directed to prospective franchisees, whose negative effect on M's expected stream of future profits is given by the per period expected loss R. Assuming, again, that all the other self-enforcement conditions are satisfied, the necessary and sufficient condition for the efficient relational contract to be self-enforcing under restricted franchising is still given by (28), whereas the necessary and sufficient conditions for the exploitative relational contract to be self-enforcing are (36) and

$$\frac{1}{r}(p_M - R) \geq \frac{1}{r} E_s \left[\pi_U \left(d^D(s), e^D(s), s\right)\right] \tag{38}$$

, which can be rewritten as

$$R \leq p_M - E_s \left[\pi_U \left(d^D(s), e^D(s), s\right)\right] \tag{39}$$

. According to condition (39), the introduction of A as a third party in the model attaches a cost to M's decision to terminate D for deviating from $d^M(s)$, and, as a result, M will terminate D only if such cost (represented by R) is low enough. It immediately follows that, if A's ability to boicott M's future business is sufficiently strong (that is, if R is high), (39) will not hold, M will not terminate D for failure to implement $d^M(s)$ and restricted franchising will not be chosen to enforce the exploitative relational contract[30].

The argument above suggests a way of empirically separating the efficiency hypothesis on the choice of restricted franchising from the exploitation hypothesis. In fact, if, *ceteris paribus*, restricted franchising is more frequently observed when R is large, then it can be said that restricted franchising tends to be chosen when it facilitates the self-enforceability of the efficient relational contract,

[30] The analysis of the three parties relational contract relies heavily on 1) A's good faith and 2) A's ability to negatively affect M's public image. In practice, there might be a tradeoff between these two attributes, in the sense that an organized A (for instance, a trade association of franchisees) would have good chances to conduct a successful campaign against M, but would also, arguably, have incentives not to limit retaliation to the cases in which M uses his termination power to exploit D.

In fact, as documented in Brickley, Dark and Weisbach (1991), at least part of existing franchisees can benefit from an *overall* weakening in franchisors' termination power and, therefore, their trade associations have incentives to denigrate franchisors even when they efficiently exert termination rights.

An incorporation of A's opportunism in the model falls beyond the scope of this article and is deferred to future research. The discussion conducted in this paragraph, therefore, must be seen as mainly suggestive and illustrative.

rather than the exploitative one. An immediate consequence is that, if one estimates the relationship between D's reneging temptation and the choice of restrcted franchising while controlling for R and the effect of R on the likelihood that restricted franchising is chosen turns out to be positive, any evidence supporting Proposition 3 is consistent with the efficiency hypothesis but not consistent with the exploitation hypothesis.

In this sense, it could be argued that the evidence presented by Arruñada, Garicano and Vázquez (2001) *weakly* supports the efficiency hypothesis, since, holding the temptation of car dealers to free ride constant, they find a positive, but statistically insignificant association between the likelihood that decision/termination rights are assigned to the manufacturers and the manufacturers' reputational capital, as proxied by the length of their presence in the market for franchises.

References

[1] AGHION, P. and TIROLE, J. (1997), "Formal And Real Authority In Organizations", *Journal Of Political Economy* 105: 1-29.

[2] ARRUÑADA, B. (2000), "The Quasi-Judicial Role of Large Retailers", *Revue d'Economie Industrielle* 92: 277-96.

[3] ARRUÑADA, B. and ANDONOVA, V. (2004), "Judges' Cognition and Market Order", *Universitat Pompeu Fabra Working Paper.*

[4] ARRUÑADA, B., GARICANO, L. and VÁZQUEZ, L. (2005), "Completing Contracts Ex-post: How Manufacturers Manage Car Dealers", *Review of Law and Economics*, forthcoming.

[5] ARRUÑADA, B., GARICANO, L. e VÀZQUEZ, L. (2001), "Contractual Allocation of Decision Rights and Incentives: The Case of Automobile Distribution", *Journal of Law, Economics and Organizations* 15: 56-73.

[6] ARRUÑADA, B. and VÁZQUEZ, L. (2003), "Organizational Choice And Environmental Change", *Universitat Pompeu Fabra Working Paper.*

[7] BAKER, G. and HUBBARD, T. (2004), "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking", *Quarterly Journal of Economics,* Nov. 2004: 1443-1479.

[8] BAKER, G., GIBBONS, R. e MURPHY, K. (2006), "Contracting For Control", *MIT Working Paper.*

[9] BAKER, G., GIBBONS, R. and MURPHY, K. (2004), "Strategic Alliances: Bridges Between "Islands Of Conscious Power"", *Marshall School of Management Working Paper.*

[10] BAKER, G., GIBBONS, R. e MURPHY, K. (2002), "Relational Contracts and the Theory of the Firm", *Quarterly Journal of Economics* 117: 39-83.

[11] BAKER, G., GIBBONS, R. e MURPHY, K. (2001), "Bringing the Market inside the Firm?", *American Economic Review* 91: 212-218.

[12] BAKER, G., GIBBONS, R. e MURPHY, K. (1999), "Informal Authority in Organizations", *Journal of Law, Economics and Organization* 15: 56-73.

[13] BRICKLEY, J., A. (2002), "Royalty Rates and Upfront Fees in Share Contracts: Evidence From Franchising", *Journal of Law, Economics and Organization* 18: 511-535.

[14] BRICKLEY, J., A. (1999), "Incentive Conflicts and Contractual Restraints: Evidence From Franchising", *Journal of Law and Economics* 42: 745-74.

[15] BRICKLEY, J., A., DARK, F., H. and WEISBACH, M., S. (1991), "The Economic Effects of Franchise Termination Laws", *Journal of Law and Economics* 34: 101-132.

[16] CROCKER, K. e REYNOLDS, K. (1993), "The Efficiency of Incomplete Contracts: an Empirical Analysis of Air Force Engine Procurement", *RAND Journal of Economics* 24: 126-46.

[17] ELFENBEIN, D. and LERNER, J. (2003), "Ownership And Control Rights In Internet Portal Alliances, 1995-1999", *RAND Journal Of Economics* 34: 356-369.

[18] EPSTEIN, R., A. (1975), "Unconscionability: a Critical Approach", *Journal of Law and Economics* 18: 293-315.

[19] FAMA, E., F. and JENSEN, M., C. (1983), "Separation of Ownership and Control", *Journal of law And Economics* 26: 301-325.

[20] GIBBONS, R. (2005), "Four Formal(izable) Theories of the Firm?", *Journal Of Economic Behaviour And Organization* 58: 200-245.

[21] GROSSMAN, S. J. and HART, O. (1987), " Vertical Integration and the Distribution of Property Rights", in RAZIN, A. and SADKA, E. (1987), "Economic Policy in Theory and Practice", St. Martin Press, New York.

[22] GROSSMAN, S. J. e HART, O. (1986), "The Costs and Benefits of Ownership: a Theory of Vertical and Lateral Integration", *Journal of Political Economy* 94: 691-719.

[23] HADFIELD, G., K. (1990), "Problematic Relations: Franchising and the Law of Incomplete Contracts", *Stanford Law Review* 42: 927-992.

[24] HALONEN, M. (2002), "Reputation And The Allocation Of Ownership", *The Economic Journal* 112: 539-558.

[25] HART, O. and HOLMSTROM, B. (2002), "A Theory of Firm Scope", *mimeo.*

[26] HART, O. and MOORE, J. (2006), "Partial Contracts", *Mimeo.*

[27] HART, O. and MOORE, J. (1999), "Foundations of Incomplete Contracts", *Review of Economic Studies* 66: 115-139.

[28] HOLMSTROM, B. (1979), "Moral Hazard and Observability", *Bell Journal of Economics* 10:74-91.

[29] HOLMSTROM, B. and MILGROM, P. (1991), "Multi-task Principal-Agent Analysis: Incentive Contracts, Asset Ownership, And Job Design", *Journal Of Law, Economics And Organization* 7: 24-52.

[30] HOLMSTROM, B. and MILGROM, P. (1994), "The Firm As An Incentive System", *American Economic Review* 84: 972-991.

[31] JANUSZEWSKI FORBES, S. and LEDERMAN, M. (2005), "Control Rights, Network Structure And Vertical Integration: Evidence From Regional Airlines", *mimeo.*

[32] KLEIN, B. (1995), "The Economics of Franchise Contracts", *Journal of Corporate Finance* 2: 9-37.

[33] KLEIN, B. (1980), "Transaction Cost Determinants of "Unfair" Contractual Arrangements", *American Economic Review* 70:356-362.

[34] KLEIN, B. e LEFFLER, K. (1981), "The Role of Market Forces in Assuring Contractual Performance", *Journal of Political Economy* 89:615-641.

[35] KLEIN, B. e MURPHY, M. (1997), "Vertical Integration as a Self-enforcing Contractual Arrangement", *American Economic Review* 87:415-420.

[36] KLEIN, B. and MURPHY, K. (1988), "Vertical Restraints As Contract Enforcement Mechanisms", *Journal Of Law And Economics* 31: 265-297.

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

[38] LAFONTAINE, F. (1992), "Agency Theory and Franchising: Some Empirical Results", *RAND Journal of Economics* 23: 263-283.

[39] LAFONTAINE, F. and SHAW, K., L. (2005), "Targeting Managerial Control: Evidence From Franchising", *Rand Journal of Economics*, forthcoming.

[40] LAZEAR, E., P. (2000), "The Future of Personnel Economics", *The Economic Journal* 110: 611-639.

[41] LEVIN, J. (2003), "Relational Incentive Contracts", *American Economic Review* 93: 835-57.

[42] MARASCO, P. G. (2005), "La Rinegoziazione e L'Intervento Del Giudice Nella Gestione Del Contratto", *Contratto e Impresa*.

[43] MASTEN, S., E. (1988), "A Legal Basis for the Firm", *Journal of Law, Economics and Organization* 4: 181-198.

[44] MASTEN, S. E. e CROCKER, K. J. (1985), "Efficient Adaptation in Long-term Contracts: Take-or-Pay Provisions for Natural Gas", *American Economic Review* 75: 1083-1093.

[45] MATOUSCHEK, N. (2004), "Ex Post Inefficiencies In A Property Rights Theory Of The Firm", *Journal Of Law, Economics And Organization* 20: 125-147.

[46] MATTHEWSON, F., G. and WINTER, R., A. (1985), "The Economics of Franchise Contracts", *Journal of Law and Economics* 28: 503-26.

[47] MEYERSSON MILGROM, E., MILGROM, P. and SINGH, R. (2006), "When Should Control Be Shared?", *Mimeo*.

[48] NASH, J., F. (1950), "The Bargaining Problem", *Econometrica* 18: 155-162.

[49] NORTON, S. W. (1988), "An Empirical Look at Franchising as an Organizational Form", *Journal of Business* 61: 197-218.

[50] PAZ-ARES, C. (1999), "La Terminación De Los Contratos De Distribución", *Revista De Derecho Mercantil*.

[51] RUBIN, P. (1978), "The Theory of the Firm and the Structure of the Franchise Contract", *Journal of Law And Economics* 21: 223-233.

[52] SIMON, H. (1951), "A Formal Theory of the Employment Relationship", *Econometrica* 19: 293-305.

[53] SCHLESINGER, P. (1992), "Poteri Unilaterali di Modificazione ("Ius Variandi") del Rapporto Contrattuale, *Giurisprudenza Commerciale* 19.1, I: 18-24.

[54] SMITH II, R., L. (1982), "Franchise Regulation: An Economic Analysis Of State Restrictions On Automobile Distribution", *Journal of Law and Economics* 25: 125-157.

[55] WICKELGREN, A., L. (2004), "On The Misuse Of The Nash Bargaining Solution In Law And Economics", *mimeo*.

[56] WILLIAMSON, O. E. (1991), "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives", *Administrative Science Quarterly* 36: 269-296.

[57] WILLIAMSON O. E. (1985), "The Economic Institutions of Capitalism", New York, NY Free Press.

[58] WILLIAMSON O. E. (1983), "Credible Commitments: Using Hostages to Support Exchange", *American Economic Review* 73: 519-540.

[59] WILLIAMSON, O. E. (1971), "The Vertical Integration of Production: Market Failure Considerations", *American Economic Review* 61: 112-123.

[60] ZANARONE, G. (2005), "Restricting Franchising: an Empirical Analysis of the Allocation of Control Rights in Car Dealership Contracts", in CAFAGGI, F. (2005), "Corporate Governance, Networks And Innovation", *CEDAM*.