Relational Contracts and Competitive Screening*

Giacomo Calzolari† Giancarlo Spagnolo‡

CEPR Discussion Paper No. 7434
This version: August, 2010

Abstract

We study the tension between competitive screening and contract enforcement where a principal trades repeatedly with one among several agents, moral hazard and adverse selection coexist, and non-contractible dimensions are governed by relational contracting. We simultaneously characterize optimal relational contracts and competitive screening policies which are interdependent. When non-contractible dimensions are important, the principal optimally restricts competitive screening to a subset of "loyal" agents, giving up performance bonuses and, when such dimensions are crucial, negotiates an indefinitely renewable contract with one agent. To enhance enforcement, explicit contract duration is also reduced. However, these policies facilitate collusion among agents, which induces an additional trade-off between reputational forces and collusion. When non-contractible dimensions are very important this last trade-off may disappear, as collusion allows more efficient enforcement of better performance.

JEL Classification Numbers: C73, D86, L14.

Keywords: Auctions, Collusion, Contract duration, Efficiency wages, Implicit contracts, Incomplete contracts, Limited enforcement, Loyalty, Multi-tasking, Negotiation, Non-contractible quality, Performance bonus, Procurement, Relational contracts, Reputation, Screening.

*This a significantly revised version of our 2004 manuscript "Reputation and Collusion in Procurement". We are grateful for their helpful comments to Pierpaolo Battigalli, Pierre-André Chiappori, Mathias Dewatripont, Nicola Doni, Sergei Izmalkov, Jonathan Levin, Claudio Mezzetti, Sebastien Mitraille, Antonio Nicolò, Yossi Spiegel, Daniel Traca, Jeanine Thal and Paola Valbonesi. We thank seminar participants at the following workshops and conferences ESSET 2005 Gerzensee, ESWC 2005 London, EARIE 2005 Porto, CEPR 2005 Competition Policy Brussels, Public Services and Management 2006 Toulouse, ESF 2006 "Designing Partnerships Between Government and The Private Sector" Bristol, II IPPC 2006 Rome, Asset meeting 2007, CSEF-IGIER 2008 Symposium Capri, Competition in Infrastructure Procurement ECARES 2009 and at the Universities of Berlin (Humbolt), Bologna, Brescia, Florence (EUI), Istanbul (Koc), Mannheim, Milan (Bocconi), Naples (Federico II and Salerno), Norwich (East-Anglia), Padua, Prague (CERGE-EI), Rome (Tor Vergata), Stockholm (KTH and Stockholm School of Economics), NBER Conference on Relational Contracts 2009, Paris Sorbonne University 2010.

†Department of Economics, University of Bologna, Italy and CEPR. E-mail: *giacomo.calzolari@unibo.it*

‡University of Rome 'Tor Vergata', SITE, EIEF & CEPR. E-mail: *giancarlo.spagnolo@uniroma2.it.*

1 Introduction

Non-contractible dimensions are present to varying measure in every economic exchange.[1] It is well known that when non-contractible tasks are important, giving agents high powered incentives on contractible tasks or letting them compete – say, in a low-price auction – may produce very inefficient outcomes, not only for a principal but in general.[2] When exchanges are recurrent, underprovision on non-contractible dimensions can be avoided by governing them with reputational/relational forces.[3] But cooperation incentives based on "the shadow of the future" are typically stronger, the larger the expected future payoffs at stake, so that enforcement may conflict with other important needs of the principal, and in particular with that of letting agents compete to screen and select the more capable. Here we study this conflict in a model where a principal trades recurrently with multiple, heterogeneous, privately-informed agents, and where non-contractible dimensions are important. We characterize the optimal relational contract, defined in the broad sense to include equilibrium choices on explicitly contracted features, on non-contractible dimensions and on the competitive screening policy. We identify situations in which tension does arise between screening and enforcement and derive the general implications of this trade off.

The importance of this tension is evident if we think, for example, of the difference between Anglo-Saxon and Japanese traditions in the management of suppliers and employees. The interaction between relational contracts and competitive screening is at the core of the comparison between Toyota's "relational" procurement policies and, say, GM's more "competitive" or "arms' length" approach (Asanuma, 1989). Like many other Japanese firms, Toyota maintains a small stable set of "highly trusted" dedicated suppliers, restricts competition for orders to them alone, cares for their profitability and rewards the best performers with a larger share of orders, while it replaces those that fail to deliver the extremely high levels of contractible and non-contractible performance required. The limits to competitive screening entail the cost of reduced screening and higher prices, at least compared to GM's more competitive selection; but this approach ensures that sufficient weight is attached to future stakes and consequently guarantees a cooperative perspective in the supply relationship. Analogously, a special long-term perspective characterizes Japanese employment relations: to ensure internal cohesion and employee "loyalty", Japanese firms have for a long-time limited competitive screening and other high-powered personnel incentives, at least by comparison with American corporations (e.g. Aoki and Dore, 1994).

Public procurement is another relevant example. Banfield (1975) drew attention early on to the likely loss of procurement quality caused by accountability rules that force public buyers to use open

[1] Reasons why some dimensions of exchanges are not explicitly contractible include complexity and prohibitive legal cost of verification; see Hart (1995) for thorough discussion and Tirole (1999) for an evaluation of the debate on contracts incompleteness.

[2] See Holmstrom and Milgrom (1991) and Manelli and Vincent (1995), among others. This contrasts with the case of full contractibility as in Bulow and Klemperer (1996).

[3] Macaulay (1963) is the classic reference.

auctions for supplier selection. Kelman (1990) followed up on this theme and on the importance of taking past performance and "the shade of the future" into proper account. He notes that private firms use open auctions much less commonly, leave higher margins to suppliers, switch suppliers less often and are more satisfied with the quality of goods and services they procure. When he became responsible for public procurement under the Clinton administration, he eased the rigidity of public procurement regulations in order to increase quality and reduce transaction costs, pushing to use data on the past performance of federal suppliers in determining qualification for future tenders, so that relational-reputational forces could start working again.[4]

Outline of the paper. We develop a dynamic model of recurrent exchange with non-contractible tasks between an infinite-lived principal and a population of heterogeneous and privately-informed infinite-lived agents. This means contemplating both moral hazard (on non-contractible dimensions) and adverse selection (on agents' type/cost). In our framework, when the principal designs the optimal relational contract he must choose: any explicit part of the contract enforceable by the court (like fixed payments/prices, contract duration, contractible performance standards); the implicit and self-enforcing incentives designed to govern costly non-contractible performance; and the screening policy, i.e. how and how often to have agents compete for selection.[5] These three aspects of the relationship are highly interdependent, so finding the optimal relational contract requires joint optimization of all three dimensions.

We first characterize the optimal relational contract and screening policy, allowing parties to exchange contractible monetary transfers (wages, participation fees, payments related to auctions' bids) as well as non-contractible ones (such as performance bonds and bonuses). The principal's temptation to renege on a promised bonus (or to withhold a bond) and then to exclude the performing agent limits the use of such incentives. Because of this, to increase the stakes of the agents and enforce non-contractible performance the buyer has to rely also on future rents increased by limiting competition among those agents pre-selected by the principal and included in a pool of eligible suppliers (at the cost of lower efficiency); and by fixed, regular transfers paid to all agents in the pool, independently of being actually supplying (similar to wages for being available to the principal).

We find that when agents compete and both unconditional transfers and performance bonuses are available, the buyer optimally chooses:

(i) discretionary performance bonuses and open auctions, when non-contractible dimensions

[4]Europe has been moving in the opposite direction, pushing for open competitive screening to minimize discrimination against foreign suppliers and to foster European market integration. This policy, made consistent with a series of EU Procurement Directives, has forced many European countries to revise procurement regimes and move towards the generalized use of rigid and open competitive procedures. A common feeling engendered by these changes, and one that is consistent with our results here, is that the main consequence has been lower-quality procurement (and an increase in transaction costs; see Europe Economics, 2006).

[5]Because we consider both contractible and non-contractible tasks, our is effectively a dynamic model with multi tasking in the sense of Holmstrom and Milgrom (1991).

are not important or there are few agents;

(ii) negotiation with a single agent on an indefinitely renewed contract conditional on non-contractible performance when the latter is crucial;

(iii) "restricted competition" in all other cases, i.e. recurrent competitive screening among a stable subset of "qualified" loyal agents (the more important non-contractible performance is, the smaller the group), under the threat of exclusion and replacement.

There are a number of important situations in which direct negotiations and discretionary transfers are not possible, either ex ante or ex post. This is often the case for large private corporations or public organizations for reasons of accountability. We find that in these circumstances the optimal approach in almost all cases is restricted competition under threat of exclusion from the set of invited bidders for poor performance.

We also show that whenever non-contractible performance is valuable, it is optimal to shorten the duration of the explicit part of the contract. Abstracting from technological aspects such as the rate of obsolescence, a shorter contract implies more frequent interaction, making it easier for the principal to get good non-contractible performance by threatening to withhold bonuses or to promptly exclude non-performing agents.[6] Furthermore, shortening contract duration also improves the efficiency of the performing agents where, as in our analysis, costs change over time.

However, an environment with few and frequently interacting agents is also the most favorable for collusive behavior between the agents. We therefore illustrate an additional and rather general trade-off between reputational forces and collusion. Longer explicit contracts – less frequent screening – and a larger pool of competing agents will deter collusion but also reduce the non-contractible performance obtainable from the competing agents. Symmetrically, shorter contracts and a smaller pool of eligible agents facilitate collusion but also the enforcement of non-contractible performance.

This new trade-off may disappoint because it seems to denote limits to the possible remedies for non-contractible tasks. However, our analysis clarifies that collusion itself may interact directly with agents' incentives for non-contractible performance. In fact, by increasing the price collusion clearly increases agents' gains from future trade with the principal which, as usual, can also be seen as the cost of being excluded for poor non-contractible performance. Hence, we show that there are circumstances in which the trade-off between non-contractible performance and collusion is only apparent. For example, when non-contractible dimensions are very important and the principal cannot use participation fees (or there are few agents), it is optimal for the principal to induce agents to collude, and the outcome is more efficient from a social point of view.

From the procurement view point this result is provocative in that it implies that a buyer is not necessarily concerned by suppliers' collusion. Yet we explain why cooperation among suppliers

[6] Some recent studies confirm this intuition even for procurement with important long-term investment needs. For railway operating companies in the UK, Affuso and Newbery (2002) show that (discretionary) investment is stimulated by shorter rather than longer contracts. Notwithstanding a standard hold-up problem associated with contract renewal that should imply the opposite, the authors suggest that frequent re-procurement with short contracts disciplines suppliers who care for future re-award of the franchise.

– common in procurement within consortia, joint-ventures and other joint bidding agreements – is indeed likely in some circumstances to result in higher performance, thus benefitting the buyer.

Related literature. Our paper contributes to the literature on efficiency wages and relational contracts that, from the pioneering contributions of Shapiro and Stiglitz (1984), Bull (1987), MacLeod and Malcomson (1989) and Baker, Gibbons and Murphy (1994) to the recent work of MacLeod (2003), Levin (2003) and Fuchs (2007), has formally analyzed the optimal design and the consequences of self-enforcing agreements for the governance of non-contractible tasks (see MacLeod 2007 for a survey).

The work closest to our own is probably that of Levin (2003), who elegantly characterizes the optimal relational contract with moral hazard and adverse selection between a buyer and a single seller, but with a major difference with respect to our analysis. Our framework allows for competition among several agents and we are thus concerned by the trade-off between competitive screening and the incentives to deliver non-contractible performance. Our work is also close to MacLeod and Malcomson (1998), which posits relational contracts between a number of principals and a larger number of competing agents, although without adverse selection and screening.

So far the literature has paid limited attention to the case of a relational contract between one principal and several privately-informed agents competing recurrently for a contract lasting for a limited span of time. Levin (2002) studies team production in relational environment but with no reference to competition or collusion. Rayo (2007) studies relational contracts between multiple agents, endogenously deriving organizational structure, but in a framework characterized by a team production problem with no adverse selection, competition or collusion. A recent paper by Board (2008) obtains a result analogous to our Proposition 2 on the optimality of limiting the number of trading partners, but in a very different model where the principal is fully informed, so competitive screening is not needed, and neither standard contractible monetary transfers (wages, prices) nor non-contractible performance bonuses (or bonds) are admitted.

Finally, our analysis is also relevant to the literature on reputation and competition, starting with the seminal work of Klein and Leffler (1981) and Shapiro (1983). These early analyses were concerned with the compatibility of "performance-assuring" reputational equilibria – requiring rents that make the effort of maintaining reputation worthwhile also with free entry in the market – but did not examine firms' competitive interaction (their incentives to steal business from one another) in detail.[7] Stiglitz (1989) raised the question of how reputation could be compatible with perfect competition, which should eliminate all future supracompetitive gains (see also Kranton 2003 and Bar-Isaac 2005). Hörner (2002) offers the first elegant answer to Stiglitz's question: in his model with heterogeneous consumers and firms, adverse selection (time-persistent costs) and moral hazard (goods' quality), high prices signal high quality and make competition compatible with reputational

[7] Shapiro (1983) first formalized the argument that higher frequency of interaction (in our paper shorter contract duration) strengthen reputational forces. See also Strausz (forthcoming) who shows that this may lead producers to increase obsolescence.

forces. In our environment, unfortunately, signalling would not work because of time-varying costs and since prices are set in a competitive winner-take-all auction (or, in negotiations, they are determined by the buyer).

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 analyzes relational contracting with competing agents, restricted competition and bilateral negotiation when the principal can use the full fledged set of instruments and also when their use is limited by the rules. Section 4 discusses the effect of agents' collusion on non-contractible performance. Section 5 discusses some extensions of the base model and concludes.

2 The base model

At any period $t = 0, 1, 2, ...$ a principal needs a task to be performed by one among $N > 1$ agents. The principal's per-period value of the task $v(q_t)$ is increasing in a costly decision $q_t (\geq 0)$ taken by the agent supplying in t, i.e. the *supplier.* The per-period cost for supplier i is $\theta_{it} + \psi(q_t)$ with $\psi(q_t)$ increasing, differentiable, concave and $\psi'(0) > 0$, $\psi(0) = 0$. The value of trade in period t with supplier i is $s(q_t) - \theta_{it}$ where $s(q_t) \equiv v(q_t) - \psi(q_t)$. Selecting an adequate measure for q_t and appropriately scaling ψ, without loss of generality we set $v(q_t) = vq_t + v_0$ with $v \geq 0$ and $v_0 \geq 0$. The time horizon is infinite, all players are risk-neutral and have a constant and common discount factor $\delta \leq 1$.

Although q_t is observable to the principal and the supplier, it is not verifiable to third parties, which makes it not-contractible. For example, q_t may be the costly effort (or specific investment) provided by an expert or an employee, or a quality feature of the procured service that cannot be specified. In the reminder we will refer to q_t as non-contractible "quality".

In our model, supply can be seen as a multi-tasking activity which, in addition to q_t, contemplates a contractible decision taken by the supplier at cost θ_{it} that is worth v_0 to the principal and generates value $v_0 - \theta_{it}$. To avoid uninteresting cases we assume $v_0 > \theta_{it}$ for any θ_{it} so that the principal never wants to discontinue supply.[8]

The principal does not know the cost θ_{it} of any agent i at any date t. Hence, he uses a competitive screening device, the *auction,* that awards at a price b_w an explicit contract requiring the winning agent to provide the specified supply for the next $x \geq 1$ periods.[9] Any standard auction format would do for our analysis; to fix ideas we refer explicitly to second price auctions. The cost of organizing an auction for the principal is $k \geq 0$.

As we will discuss at length, the principal may want to restrict participation, in which case he forms a *pool of* $n (\leq N)$ *competing agents* and may even restrict n to 1, thus making take-it-or-

[8] Our aim is to study simultaneous interaction between screening and enforcement, hence we will not consider multiple sourcing.

[9] To avoid integers problems, we treat x as a continuous variable.

leave-it offers to a single agent (hereafter, negotiation). Since they are ex-ante identical, the agents admitted in the pool are chosen randomly when $n < N$. We accordingly define the following.

Definition 1 (Screening modes) Open competition*: the principal allows all agents to participate and sets* $n = N$; Restricted competition: *the principal restricts participation setting* $n < N$, *which also includes the limit case of* Negotiation *when* $n = 1$.

The principal and the agents may be able to exchange other monetary transfers besides price b_w for the contract. In particular, before competition takes place and agents learn their costs, they may exchange *an ex-ante transfer* $w \in \Re$ (e.g. a fee $w > 0$ for participation in the selection process or a fixed wage $w < 0$ to compensate agents for remaining available to the principal), and, at the end of a contract, they may also exchange an *ex-post transfer* $B \in \Re$ (e.g. a discretionary bonus that the principal may pay or a performance bond posted by the supplier at the beginning of the contract that the principal may decide discretionally not to return). We will also analyze the possibility that some of these transfers are not feasible – say, for institutional reasons.

As in Levin (2003), $\theta_{it} \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$ is independently and identically distributed with density $f(\theta_{it})(> 0)$ and $\theta_e \equiv E(\theta)$ (see Section 5 for a discussion on IID costs).[10] To simplify exposition we assume agents are fully informed (but the principal is uninformed). This may be justified in environments with frequently interacting agents that know each other well (quite a common circumstance in procurement, for example). This assumption is immaterial except for the results in Section 4, in which case we show that the trade-offs underlying our results persist qualitatively even with asymmetrically informed agents.

Although quality is not contractible, the principal and the agents may still profit from ongoing interactions and reach some agreement, the *relational contract* C, about how they are going to behave in the future in and out of equilibrium. C is self-enforcing if it describes a perfect public equilibrium of the repeated game (Fudenberg, Levine and Maskin, 1994). In the following we refer to the explicit contract lasting x periods simply as the "*contract*" and distinguish it from the implicit (self enforcing) part of the relational contract C. In our environment any equilibrium allocation supported by non-stationary strategies can also be supported by stationary strategies since incentives can be provided to all parties through immediate compensation without modifying future continuation payoffs (MacLeod and Malcomson, 1989 and Levin, 2003). Hence, we concentrate on stationary contracts where behavior does not change along the equilibrium path of play.

The principal thus sets a relational contract that, in addition to the contractible terms (e.g. the selection rules, contract length x and participation fee w), also specifies a required performance level q that the supplier must deliver in all x periods (that q is the same for x periods is without loss), the maximum number n of competing agents and possibly also the discretionary bonus B that the principal may decide to pay at the end of the contract.[11] Upon a deviation at date t,

[10] That cost is drawn anew also in any of the x periods of a given procurement contract is immaterial for all results.

[11] Although not always realistic, assuming n contractible would strengthen our results. As in MacLeod and Mal-

either by the date-t supplier or by the principal, C prescribes how parties will behave in the future. Following MacLeod and Malcomson (1989, 1998), we posit that whenever there is a deviation the two separate forever (the most severe punishment, Abreu, 1988), and C prescribes that if the relationship between agent i and the principal falls apart, whenever possible the latter replaces i with another agent from the pool of $N-n$ previously excluded agents (i.e. if $n<N$). However, with monetary transfers one can always determine a self-enforcing relational contract $\tilde{C}$ supporting the same equilibrium payoffs as C but that does not require separation and is renegotiation proof (Levin, 2003).[12]

If the relationship between the principal and an agent falls apart, none of the other players knows the identity of the deviator. Consistently and as in most papers on relational contracts, we rule out external reputation effects, which we discuss in Section 5.[13]

The **timing of the game** is as follows.

$\mathbf{t=-1}$**:** *The principal sets the relational contract C.*

$\mathbf{t=0}$**:** *n agents are randomly chosen and decide whether to participate in the infinite repetition of the following* ***"auction" stage-game****:*

– *At time $t_1 \geq 0$ the auction selects a unique supplier and the principal incurs the cost k;*

– *At any period $t \in \{t_1, ..., t_1+x-1\}$ the supplier procures q;*

– *At time t_1+x, if neither party has cheated a new stage game begins, otherwise the principal decides whether or not to exclude the current supplier. Then, a new stage game begins.*

At date t the cost for any agent i to supply for x periods will be $\theta_{it}+\theta(x)+c(q,x)$ with $\theta(x) \equiv \theta_e(\delta-\delta^x)/(1-\delta)$, and a cost of quality $c(q,x) \equiv \psi(q)(1-\delta^x)/(1-\delta)$. We define $\theta'(n)$ as the smallest element in $(\theta_{1t}, ..., \theta_{nt})$ and, to simplify notation, we proceed as if the event $\theta_{it}=\theta_{jt}$ for some $i \neq j$ has zero measure at any t.

Before proceeding let us conclude this section by briefly illustrating the benchmark of *"zero-quality" equilibria* in which at any t the supplier provides $q_t=0$ and the principal requires $q_t=0$ (and clearly sets $B=0$) getting a payoff

$$V_0 = v_0 \frac{1}{1-\delta} - [nw + E(b_w) + k] \frac{1}{1-\delta^x}.$$

In any auction with $n>1$ competing agents, the most efficient one wins, gets a price $b_w = \theta''(n)+\theta(x)$ where $\theta''(n)$ is the (first-period) cost of the second most efficient agent, and a profit

comson (1989), bonds can be seen as negative bonuses. Bonuses and bonds exchanged between the principal and any agent that is not the current supplier are irrelevant.

[12] With $\tilde{C}$, upon a deviation any bilateral relationship is kept in being but the bilateral surplus produced by the date-t supplier and the principal is reallocated with appropriate transfers. The cheating party is punished to the outside-option and the other party obtains all the remaining bilateral surplus. This "Pareto-perfect" continuation is subgame perfect and guarantees that no surplus is wasted.

[13] One notable exception on external reputation is Levin (2002). Here we allow that if any of the n agent is directly harmed (i.e. gets a smaller payoff) by a deviation of the principal, both when i is or is not the current supplier, then the deviation is detected by agent i.

$w+b_w-\theta'(n)-\theta(x)$. Hence, at any auction an agent may expect to earn a *profit* $w+\beta(n)\pi(n)$ where $\pi(n)$ is agent's expected *informational rent* $\pi(n) \equiv E[\theta''(n)-\theta'(n)]$ and $\beta(n)$ is the probability of being the most efficient among the n agents.[14] The principal then extracts $\pi(n)$ by optimally setting w which is thus a participation fee for all agents in the pool. Since the expected cost $E(\theta'(n))+\theta(x)$ of procuring a contract lasting x periods is decreasing in n, the principal prefers inviting all agents to bid. As for contract duration, the principal then faces a simple trade-off. He compares the cost k of organizing a new auction with the associated price reduction, i.e. the difference between the expected cost θ_e for one more period under the same contract and the cost $E[\theta'(n)]$ if he terminates the contract and runs a new auction that selects the most efficient agent for the next period.

Lemma 1 (Zero-quality equilibria) *The optimal relational contract with zero-quality C_0 prescribes $B_0 = 0$, $n_0 = N$ and $x_0 = 1$ if $\theta_e - E\left[\theta'(N)\right] \geq k$, $x_0 = \infty$ otherwise.*

Notice that, since the cost saving $\theta_e - E[\theta'(n)]$ from an auction with n agents is increasing in n, stronger competition induces (weakly) shorter contracts: n_0 and x_0 are substitutes. Although the principal may select a single seller and keep it forever so as to save the cost k of auctions, this decision is always dominated by inviting all agents. We denote the principal's maximal payoff in a zero-quality equilibrium as V_0^*. Also notice that if the participation fee w were not feasible, the principal could not extract the informational rent $\pi(n)$, but the only difference would be that the actual first period cost would be $E[\theta''(n)]$ instead of $E[\theta'(n)]$ and the results in Lemma 1 would hold with this difference.

Finally, were quality fully contractible, the optimal contract C_{FB} would reach the *first best* with the same contract length x_{FB} and the same number of competing agents n_{FB} as in C_0 but a quality q_{FB} that maximizes $v(q)-\psi(q)$.

3 Non-contractible quality, competition and negotiation

In any relational contract C_n with n competing agents that prescribes a strictly positive quality q, agents may be tempted to cheat and save on quality costs with two types of deviation. First, the supplier that has won an auction may decide to deliver zero quality on the current contract. Second, anticipating that he will cheat on quality in case of being awarded the contract and therefore expecting lower costs than his competitors, an agent may bid aggressively and win even if it is not the most efficient. The principal can control both types of deviation by guaranteeing participating agents future profits greater than what they can get by cheating. In any future auction, any of the n agents may expect to benefit (i) from simple participation through the transfer w, and (ii) from

[14] In case $n = 1$, we would have $\pi(1) = r - \theta_e - \theta(x)$ where r is the highest admissible bid. In this case, since the principal forgoes screening, w and r are equivalent transfers and, to limit notation, we explicitly consider only w. Hence, the expected profit for a given contract is simply defined as $\pi(1) = w - \theta_e - \theta(x)$.

winning the auction and thus getting the (expected) price $E(b_w)$ and possibly receiving the bonus B. On the other hand, in case of contract award any agent expects a cost $E(\theta'(n)) + \theta(x) + c(q,x)$, so that the incentives to participate and to deliver quality are characterized as follows.

Lemma 2 (Firms' incentives) *Consider any relational contract C_n prescribing $q > 0$. It induces the n agents to participate, to bid with the intention of delivering q and to actually deliver q if*

$$w + \beta(n)\pi(n) + B\ \mathcal{I}_{n=1} \geq 0, \tag{1}$$

$$[w + \beta(n)\pi(n) + B\ \mathcal{I}_{n=1}]\,\frac{\delta^x}{1-\delta^x} \geq c(q,x) - B. \tag{2}$$

Participation constraint (1) guarantees agents non-negative expected profits at any auction. When there is competition (i.e. $n > 1$), agents are induced to lower their bid by any transfer that the principal may want to pay exclusively to the winner, in addition to b_w. Hence, ultimately the expected profit $w + \beta(n)\pi(n)$ does not depend on B, unless the principal prefers avoiding competition and contracts with a single agent, setting $n = 1$.[15] If agent's incentive compatibility constraint (2) is satisfied, any supplier who has won any given auction prefers not to cheat in the current contract since the present value of expected future profits is greater than the cost saving $c(q,x)$ from cheating on quality but then losing both the bonus B on the current contract and all future profits. Furthermore, no agent i at any auction ever plans to cheat even if by doing so it can reduce its bid and so certainly win contract award. Indeed, with this deviation no agent (independently of its efficiency θ_{it}) can ever gain more than the term on the right hand side of (2) less $\theta_{it} - \theta''(n) \geq 0$.

Condition (2) shows that both a larger w and a larger B make cheating more costly for agents. Furthermore, a larger number of competitors n reduces both the probability of winning $\beta(n)$ and the informational rent $\pi(n)$. Similarly, a longer contract, i.e. higher x, makes rents more distant in the future and increases the cost of the current contract. Hence, both stronger competition and longer contracts have an adverse effect on agents' incentives to perform.

If the principal can use it credibly, the bonus is a very effective tool for controlling agents' incentives since it can be retained at the end of the contract, whereas the fixed fee w is paid in advance. This desirable property of B has been illustrated by several papers on relational contracting (e.g. MacLeod and Malcomson, 1989 and Baker et al., 1994). The framework proposed in this model illustrates some interesting and hitherto unexplored features of the bonus that are specific to relational contracting with competing agents. First, as explained, the bonus is never actually paid and so costs nothing to the principal since competition "sweeps it away" by reducing the payment b_w precisely by the amount of B (as it would be to any fixed transfer paid exclusively to the supplier in addition to b_w). Second, inducing incentives with a larger B may let the principal reduce the fixed fee w that must instead be paid to *all* the n bidding agents.

[15] In this case, however, we will show that the only credible bonus is $B = 0$.

However, the possibility of using the bonus may be limited by the principal's incentives. To see this, consider the principal's expected present value payoff[16]

$$V_n \equiv s(q)\frac{1}{1-\delta} - [nw + E(b_w) + B + k\mathcal{I}_{n>1}]\frac{1}{1-\delta^x}.$$

At the end of any contract (having already paid b_w and w), the principal may renege on the bonus. In this case, the relationship with the current supplier breaks down and the principal is left with the outside option $\underline{V}_n$, i.e. the expected payoff when the current supplier cannot be employed anymore.

We can then state the following.

Lemma 3 (Buyer's incentives) *Any relational contract C_n prescribing $q > 0$ is incentive-compatible for the principal if*

$$(V_n - \underline{V}_n)\delta \geq B, \tag{3}$$

and is individually rational if

$$V_n \geq V_0^*. \tag{4}$$

With restricted competition, $B = 0$ in any relational contract C_n that satisfies principal's incentive compatibility (3). With open competition there exist relational contracts C_n that satisfy (3) with $B > 0$.

When the principal sets $n < N$, thus instituting restricted competition, he can always renege on the bonus and replace the current supplier at no cost with another agent among those $N-n$ previously excluded from the pool. This guarantees a continuation payoff $\underline{V}_n$ that is at least V_n, so that the unique credible bonus is zero.[17] If instead the principal admits all agents in open competition, i.e. $n = N$, then a positive bonus may be credible. In fact, starting with all N agents, if now the principal reneges the bonus, he cannot replace the current supplier and his outside option $\underline{V}_N$ is determined by the optimal relational contract he designs with only $N-1$ agents, i.e. V_{N-1}.[18]. Finally, the principal's participation constraint (4) shows that he can always secure the zero-quality payoff.

We can now characterize the principal's optimal contract C_n^*, which maximizes V_n subject to incentive compatibility and the participation constraints of all players, i.e. (2)-(4). Since V_n is

[16] Notice that if $n = 1$ the principal need not select the supplier so that he does not pay the cost k of organizing the search process.

[17] This observation is related to efficiency-wage equilibria in the labor market where workers (in our model, agents) on the "long side" and firms (in our model, the principal) are on the "short side" of the market (see for example MacLeod and Malcomson, 1998). In these cases too, bonuses are not credible.

[18] As explained in the proof, upon a deviation, none of the $N-1$ remaining agents are negatively affected by the principal's deviation. Furthermore, after a deviation from a contract C_N, the possibility to set $n < N-1$ is dominated for the principal by contracting instead with all $N-1$ agents.

decreasing in the fee, w is optimally set such that (2) and (1) are binding, i.e.

$$w = \frac{1-\delta^x}{\delta^x}\max\{c(q,x)-B,0\} - \beta(n)\pi(n) - B\ \mathcal{I}_{n=1}. \tag{5}$$

This shows that whenever the principal can control agents' incentives on quality by setting B equal to the cost $c(q,x)$, then w allows him to extract all the expected rents of bidding agents, as in the zero-quality equilibrium. In this case, w is a participation fee. If instead $B < c(q,x)$, then w may become negative, thus constituting a transfer from the principal to the competing agents to control their expected profits and incentives. In this case, the transfer w can be seen as an employment contract under which the principal pays in order to have the n agents always available.

Substituting w and $E(b_w)$, the principal equivalently maximizes

$$V_n = s(q)\frac{1}{1-\delta} - [E(\theta'(n)) + \theta_e(x) + k\ \mathcal{I}_{n>1}]\frac{1}{1-\delta^x} - \frac{n}{\delta^x}\max\{c(q,x)-B,0\}, \tag{6}$$

subject to (3) and (4). This expression for V_n shows that, unless the bonus B exactly matches the cost $c(q,x)$, quality is distorted for incentive reasons, and the distortion increases with the number of competing agents n (the last term in (6)). The principal's problem is that if $B < c(q,x)$ then, in order to give the right incentives to any prospective supplier to procure quality, he must increase the transfer w by the first term on the right hand side of (5), and since w is paid to all competing agents (see the discussion above), the cost of providing incentives for quality is finally the last term in V_n.

Proposition 1 (Restricted and open competition) ***(i) Restricted competition.*** *The optimal relational contract C_n^* employing $n < N$ agents is such that: q_n^*, x_n^* and n^* are distorted downward with respect to the first best; making quality more important (i.e. larger v) induces a larger q_n^* but smaller x_n^* and n^*; substitutability between n^* and x_n^* is stronger than in first best; if negotiation is optimal (i.e. $n^* = 1$) then $x_1^* = 1$.*

(ii) Open competition. *There exists a decreasing function $\underline{N}(v) \geq 0$, with $\underline{N}(0) = \infty$ and $\underline{N}(v) = 0$ for any $v \geq \bar{v}$ (> 0), such that if the number of agents is sufficiently small, i.e. $N \leq \underline{N}(v)$, then the optimal relational contract with N agents is fully efficient, i.e. $C_N^* = C_{FB}$; otherwise q_N^* and x_N^* are distorted downward with respect to the first best and decreasing in N with $\lim_{N\to\infty} q_N^* = 0$ and $\lim_{N\to\infty} x_N^* = 1$.*

Under restricted competition the relational contract always contemplates distortions since to provide incentives for the competing agents, the principal must lower the quality demanded to the supplier, as is indicated by the last term in (6). How much the principal wants to restrict competition depends on a trade-off on n. On one hand, a larger n reduces the cost of procurement via a smaller expected cost $E[\theta'(n)]$. On the other hand, a larger n reduces agents' expected rents thus making the incentive-compatibility constraint of the supplier harder and more costly to satisfy (as shown in (2) and in the last term of (6)). This trade-off may also lead the principal to set

$n = 1$ thus renewing the contract with the same supplier (unless one of the two cheats) in which case it is optimal to set the shortest possible contract duration since with negotiation there are no costs k for the search process. The optimal contract length for $n > 1$, in addition to what we have seen for the zero-quality equilibria (Lemma 1), also accounts for the fact that a longer contract now increases the cost of quality $c(q, x)$ and again implies higher incentive costs to control quality. The contract, then, is shortened for incentive reasons. It is also worth noticing that if the principal wants to procure higher quality (e.g. since v is larger), this increases the distortionary (last) term in V_n (i.e. $n/\delta^x\ c(q, x)$) and, consequently, the principal will optimally further reduce both x and n.

Consider now open competition. In this case the principal can credibly use the bonus and will in fact increase B as much as possible. What is more, and in contrast to the standard relational contracting models with a single agent, under competition the principal always optimally sets B high enough that the agent's incentive compatibility constraint (2) is satisfied; that is, he raises B up to $c(q, x)$. With competing agents a larger B makes incentive compatibility of the principal (3) "easier" to satisfy, since the bonus is never actually paid, and it makes it possible to lower the fee w paid to all N agents. As a consequence, the principal optimally sets $B = c(q, x)$ (an even higher B has no effect on V_N and is "costly" in terms of constraint (3)). This opens up the possibility that the optimal relational contract may be fully efficient, since the distortion in V_N is eliminated. This possibility ultimately depends on the number of agents N. To see this notice that when quality and contract length are as in the first best, the difference $V_N - V_{N-1}$ in the left hand side of (3) simply reflects the cost-efficiency gain for the principal by having one additional agent instead of contracting with $N - 1$ agents, i.e. $E[\theta'(N-1)] - E[\theta'(N)]$ which is decreasing in N. Hence, if N is sufficiently small, then it may well be the case that $(V_N - V_{N-1})\delta$ is larger than $c(q^{FB}, x^{FB})$ and the relational contract is fully efficient. But if N is large, the present value of the gain from one additional agent is less than the cost of procuring the efficient level of quality. In this case, the principal's incentive constraint (3) is not satisfied and he must reduce both q and x. Furthermore, since $V_N - V_{N-1}$ is bounded from above by $E[\theta'(N-1)] - E[\theta'(N)]$ which is decreasing in N, quality and contract duration are both decreasing in N and tend respectively to zero and to $x = 1$ as N becomes larger and larger.

In light of Proposition 1, what, in the end, is the best screening mode? If the structure of the supply market is such that there are only few agents able to perform the task (i.e. N small), then open competition may allow the principal to reach the first best and any other screening mode is dominated. However, we know that when the number of competing agents is large enough, open competition does not deliver the first best and the optimal screening mode depends on two different distortions.

Proposition 2 (Optimal procurement mode) *There exists a decreasing function $\overline{N}(v) \geq \underline{N}(v)$, with $\overline{N}(v) - \underline{N}(v)$ increasing in v and $\overline{N}(0) = \underline{N}(0)$, such that for $N \leq \underline{N}(v)$ open competition is optimal while for $N \geq \overline{N}(v)$ restricted competition is optimal with a number of competing agents*

$n^*(v)$ that is (weakly) decreasing in v and in the limit $n^*(v) = 1$ (negotiation). In intermediate cases, i.e. $N \in (\underline{N}(v), \overline{N}(v))$ either open or restricted competition may be optimal.

If quality is not at all important for the principal, i.e. $v = 0$, we already know that maximal competition among agents – open competition – is the best choice for the principal. This is also the case for positive but small v, if there are few competing agents, since we know that in this case open competition eliminates any distortion. But where quality is very important for the principal, i.e. v is sufficiently large, then he must control suppliers' incentives to deliver q by limiting competition. Indeed, open competition entails a strong quality distortion since the bonus is strictly limited by the principal's incentive constraint and rents to the agents are small. In this case, then, the principal prefers restricted competition so as to control the number of agents n and their rents; this can lead to negotiation with $n = 1$ for very high value of quality v.[19]

3.1 Procurement with limited instruments

As noted in the Introduction, often principals cannot use discretionary bonuses and/or participation fees w. It becomes therefore important to verify whether the foregoing results are robust to restriction of the instruments available to the principal.

Corollary 1 (Screening with limited instruments) *If the principal cannot use the bonus B and/or the fixed transfers w, then the optimal number of competing agents n is further reduced, and restricted competition is optimal for a larger set of parameters.*

In this case, enforceable quality is lower, the less heterogeneous the agents are.

If the bonus cannot be used, the only difference between restricted and open competition is the larger number of competing agents in the latter. If the fixed fee w cannot be used either, then the only way to govern supplier's incentives to provide quality is through the supplier's expected informational rent $\beta(n)\pi(n)$, because agent's incentive-compatibility constraint (2) becomes

$$\beta(n)\pi(n) \geq \frac{1-\delta^x}{\delta^x} c(q, x). \tag{7}$$

Since the expected informational rent $\beta(n)\pi(n)$ is decreasing in n (both terms decrease with n), the only way to increase q while ensuring incentive compatibility is to further restrict the number of competing agents, with the highest enforceable quality obtained negotiating with a single agent.[20] When w cannot be used, condition (7) implicitly defines the *maximum enforceable quality* $q(n, x)$ decreasing in n and x – that the principal can ask for any n and x – so that any q is incentive compatible if and only if $q \leq q(n, x)$.

[19] Clearly, a higher cost k of organizing auctions makes contract renewal relatively more desirable.

[20] As shown in the Appendix, an analogous argument applies where the principal can use B but not w.

When the set of instruments available to the principal is limited and he must rely on the expected informational rent to govern quality, then a low variability of agent's cost θ reduces the level of enforceable quality. At the limit, when agents are homogeneous and $\pi(n) = 0$ for any $n > 1$, if quality is important the principal should then procure with private negotiation, since this is the only way to guarantee the agent some future rents.

4 Reputation and collusion

We have shown that where quality is important for the principal (i.e. v is high), to control for non-contractible quality may require shortening contract length x (Proposition 1) and reducing the number of n agents competing (Proposition 2 and Corollary 1). It is well known that both courses of action effectively tend to foster collusion among agents. The results of Section 3 therefore indicate that in controlling for quality the principal may risk inducing agents to collude.

To take the possibility of agents's collusion into account, we now denote with $\tilde{\pi}(n)$ the supplier's expected rent when there is a cartel among the n agents. In this case, the most efficient agent is awarded the contract and all the others in the pool of n agents either refrain from bidding or submit losing bids; that is, collusion takes the form of bid rotation.[21]

First we let the principal use the full set of instruments; afterwards, as in Section 3.1, we extend the discussion to the effects of a limitation on them. Since collusion among agents affects not only their payoffs but also their incentives to deviate, we now explicitly let the principal devise a relational contract anticipating that agents will be induced to collude.

For collusion to be sustainable at any auction, the second most efficient agent (i.e. the one with the greatest incentive to cheat) must not prefer to undercut the most efficient. If this agent does not deviate, he can expect the future collusive profits. Otherwise, by deviating he gets an immediate gain $D \geq 0$ but then collusion breaks down and all agents will compete from then on. Clearly, a deviating agent may also consider the possibility of cheating on quality. Hence, collusion with the requested quality q is viable if the following incentive compatibility constraint is verified,

$$[w + B + \beta(n)\tilde{\pi}(n)] \frac{\delta^x}{1-\delta^x} \geq D + \max\{B + [w^* + \beta(n)\pi^*(n)] \frac{\delta^{x^*}}{1-\delta^{x^*}}, c(q, x)\} \qquad (8)$$

where the variables indicated with an asterisk refer to the optimal relational contract derived in Section 3, since if collusion brakes down the principal realizes that agents have reverted to competition and adjusts the relational contract accordingly.[22]

[21] We do not consider partial collusion involving fewer than n agents. To simplify exposition, we also assume collusion is incentive-compatible for any realization of costs. Alternatively, for a sufficiently low realization of costs the cartel may contemplate temporary reversion to competitive bids (as in Rotemberg and Saloner, 1986), but this would not qualitatively alter our results.

[22] The winning bid tells the principal whether collusion has broken down or not. Assuming that the relational contract remains unchanged after cooperation breaks down may fail to satisfy subgame perfection.

Now, when collusion is in place and it is stable (i.e. condition (8) is satisfied), the current supplier prefers not to cheat on quality if

$$[w + B + \beta(n)\tilde{\pi}(n)]\,\frac{\delta^x}{1-\delta^x} \geq c(q,x) - B. \tag{9}$$

Notice that with collusion the supplier does earn the bonus B and the principal does pay it since collusion means that agents need not reduce their bids by B.

Since the left hand side of the incentive-compatibility constraint (8) for collusion is decreasing and the right hand side is (weakly) increasing in x, then by reducing contract length x the incentives to collude are strengthened. Similarly, a smaller n increases the probability $\beta(n)$ of being the most efficient agent and, since also $\tilde{\pi}(n) \geq \pi^*(n)$, a smaller pool of potential suppliers n makes collusion stronger.[23] As anticipated, when the principal wants to raise quality, he also increases the scope for collusion and its stability. This line of reasoning leads immediately to the following.

Proposition 3 (Reputation and collusion) *There is a trade-off between reputation and collusion: setting x and n lower in order to increase q via greater future profits for agents, the principal tends to induce collusion among agents.*

This is a quite general trade-off that is clearly relevant in many other frameworks that share the ingredients of potential competition among agents and the need to give them the incentives to perform non-contractible tasks.

How does the principal address this trade-off between non-contractible quality and collusion? In this light, what is the optimal screening mode? We begin answering these questions with the following Lemma.

Lemma 4 (Quality with collusion) *If quality q is enforceable with competing agents, it is so also with colluding agents.*

This is a simple consequence of the fact that stability of collusion (i.e. constraint (8)) implies that the expected profit with collusion is greater than with competition. Hence, any supplier when deciding whether or not to deliver q knows that the cost of cheating on quality is greater than under competition, because the profits at stake are larger. It is important to reiterate that, as stated in the Introduction, what we here call "collusion" can be taken to stand as any cooperative agreement among agents such as self-sustaining consortia and other forms of joint bidding like joint ventures. In this case, the principal would equivalently negotiate with a single consortium composed with the same number of agents and incentive compatibility could be seen as internal incentives for the stability of the consortium regardless of any legal obligation among agreeing partners. It is also

[23] For simplicity of exposition we omit the fact that, for given collusive bid, a smaller n also reduces D (since the second most efficient agent becomes more efficient), and the fact that $\pi(n) - \pi^*(n)$ is also decreasing in n. Both these effects strengthen our argument.

worth mentioning that Lemma 4 also holds independently of the principal's ability to use B or w. Indeed, where these instruments are not available, the maximum enforceable quality is $q(n, x)$ (as defined by agents' incentive compatibility constraint (7)) which is higher the higher is the expected informational rent.

We can now state the following.

Proposition 4 (Optimality of collusion) *Assume the fix fee w is not practicable for the principal. Then, for a high enough value of quality v, inducing collusion among agents is optimal when either (a) the principal cannot procure via negotiation, or (b) he also cares about procurement efficiency in addition to V_n.*

Where w is feasible, collusion is optimal for the principal if it is associated with a sufficiently high B and N is not too high.

When quality is important to the principal, limiting competition may become desirable above and beyond the principles of Proposition 1. Not only may the principal optimally want to restrict the number of competing agents, but he can improve quality further by inducing agents to cooperate in consortia or cartels, because with collusion the agents' expected profits increase, which ultimately makes better quality attainable.

This is certainly the case if the fixed fee w cannot be used, since otherwise the principal could directly increase agents' rents and incentives directly. Clearly, if the principal can restrict competition to the case of single-party negotiation, this gives the largest possible rent to the single supplier. However, we know that there are cases in which a buyer cannot restrict attention to a single supplier (as it is often the case in public procurement), so that where quality is important enough, inducing agents to collude turns out to be optimal.[24] Furthermore, the principal may also have a general concern for efficiency in production (it is often the case, again, in public procurement). If this is the case, then the higher rent that could be given to the single supplier under negotiation must clearly be compared with the efficiency reduction of the single agent as compared with $n > 1$ colluding agents. The balance between high quality and high efficiency ultimately induces the principal interested in both to choose cartel/consortium formation.

Consider now the case in which it is possible to use the transfer w. The benefits of collusion are now reduced, because the principal can now control agents' profits and also quality directly with w. Now, however, when we compare competition with collusion then the fact that under collusion B is effectively paid allows the transfer w to all the agents to be reduced accordingly. Hence, if $B > 0$ and N is not too large (otherwise we know from the previous section that B must be low to control the principal's incentives), then the positive effect of B is sufficiently large with collusion so that the principal will prefer it to competition whenever competition with N agents was optimal in the previous Section.

[24] In most countries' small-scale public procurement and that of international organizations like the United Nations, accountability rules require getting competitive offers from some minimum number of potential suppliers.

Our analysis illustrates some benefits of consortia and joint bidding among otherwise rival firms in procurement (in terms of enforceable quality) but also their possible limitations by accounting explicitly for the requirement that they be stable. Furthermore, in the cases in which consortia are barred by law, inducing agents to collude implicitly is a way of recovering the benefits of cooperation in terms of higher quality. In other cases, consortia are admissible and do not need to be self-enforcing (in that they are governed by explicit contracts). If this is the case and collusion among agents is optimal in terms of our previous analysis, a consortium is then better still, since it guarantees higher rents to the firms and equal efficiency with collusion.

5 Robustness and Extensions

We have illustrated how a principal addresses the tension between screening and enforcement when heterogeneous agents are available to provide a good or a task of non-contractible quality. Our framework is simple, but rich enough to deliver insights on the degree of competition, the structure of compensation with bonus and transfers and the contract length that are optimal for the principal. We have shown that when the quality of performance is important, the principal tends to prefer shorter contracts and limited competition that may also lead agents to cooperate (collude) in the interest of the principal.

Our results are robust to a number of complications and the model may accommodate some interesting extensions.

Alternative mechanisms. In the foregoing we have posited a mix of ex-ante and ex-post transfers whereby the principal controls agents' future profits and there are no other types of transfer available to increase the suppliers' profits, and then control performance, in addition to the fee w paid to all bidding agents, the price to the winner b_w and the bonus B. Our approach is thus general enough to deliver results robust to more complex mechanisms.

Although we did not explicitly consider the possibility of the principal setting a minimum price l (so that admissible bids must be $b_i \geq l$), with collusion or negotiation a minimum price is irrelevant. With competing firms it would be detrimental for the principal where he can use the transfer w (a fortiori if he cares for efficiency), because with a minimum price the supplier is not systematically the most efficient one (all agents with costs larger than l bid the same price l and thus have equal chances of winning the contract). When instead w is not available, a larger l may substitute for a reduction in n because it allows to increase agents' expected rent, although at a cost of efficiency reduction. Hence, our analysis sheds light on the observation that minimum prices are a common practice in public procurement when the purchaser cannot restrict the number of the potential suppliers; but they are rare in private procurement where the principal can limit competition.

Finally, the principal may decide to rank agents' offers by a scoring rule that is a function of the bid (i.e. the price) and the performance promised. In this case, instead of relying on simple quality as above, the principal would exclude a supplier that failed to generate the promised score ex-post.[25] However, this form of competition with bid-quality offers and a scoring rule is irrelevant as long as there is no heterogeneity among agents on the cost of the performance, as in our model.

Cartels, joint bidding and sub-contracting. Cartels, consortia and joint bidding may act even more efficiently than we have supposed above. They can boost efficiency by delegating (or subcontracting) production to the agent that is the most efficient in any period. In this case, clearly, the optimality of procuring with a cartel or a consortium increases with respect to both competition and negotiation. For simplicity we also considered only consortia or cartels that comprise all the n agents admitted at the bidding stage, but this is not the sole possibility. The principal may induce partial cartellization, for example allowing consortia of at most two agents, so that the incentives to deliver non-contractible performance, the price and the level of efficiency could be further "fine tuned" to the principal's interest.

Another element that we have not explicitly addressed is that a cartel or a consortium may also help agents to monitor the principal's behavior. The common organization may allow all the cooperating agents to verify quality and then check any deviation by the principal. This possibility is known in the literature as multilateral relational contracting (see Levin, 2002), to be contrasted with bilateral relational contracting that we analyze here. Since with multilateral relations the cartel pools the reaction of many agents to deviation by the principal, this may discipline the latter and also permit the attainment of higher non-contractible performance.

Finally, sub-contracting is a common practice, for example in procurement, and its effects may turn out to be complex: since the responsibility for the performance may remain with the lead contractor, the incentives may be diluted. We plan to investigate this interesting relationship between sub-contracting and non-contractible performance in a future work.

Alternative information structures. Though here we assume fully informed agents, the drivers of our simple trade-offs also hold in a more complex environment with privately-informed agents. Indeed, the analysis in Sections 2 and 3 can be extended to asymmetric information almost without modification (applying the revenue equivalence theorem all standard auction formats would remain equivalent for the principal).[26] Concerning colluding agents, significant complexities have been discussed by the literature on collusion with repeated auctions (see Skrzypacz and Hopenhayn 2004 and Blume and Heidhues 2006); and as the cartel members are privately-informed, the efficiency

[25] The possibility of using scoring rules may be limited by the fact that the assignment of the contract (i.e. a contractible dimension) would be determined in part by non-contractible dimensions. In the case of public procurement, for example, this may not be viable.

[26] Since with privately informed agents "Maskin mechanisms" are not available (i.e. mechanisms in which the principal would punish costs reports by agents that are not consistent), we have not considered this possibility in the previous pages either.

properties of the cartel would be weakened. However, what matters for our results on collusion is the simple fact that the latter is incentive-compatible. Since this necessarily implies that (equilibrium) expected profits with collusion are greater than with competition, ultimately, enforceable performance is higher than with competing agents. Furthermore, the comparison between several colluding agents and negotiation is also qualitatively unaffected since it is clear that, on average, the supplier selected out of the many agents is more efficient than the single agent with negotiation.

In our analysis we followed Levine (2003) and other previous work on relational contracts with asymmetric information in assuming that agents' efficiency is IID. An interesting avenue for future research is introducing cost persistency in our model. With cost persistence, the principal would learn from auctions and the cost of dismissing a cheating but efficient agent would be higher than in our environment. Setting aside the complications of such a model, in this case exclusion could be less of a deterrent for efficient agents suggesting an intrinsic trade-off between efficiency and performance. On one hand, less efficient agents would be aware that they can be readily discarded and replaced and this provides the right incentives. On the other hand, more efficient agents know that the principal would be reluctant to discard them and so will be less disciplined in providing high non-contractible performance. The analysis of this novel and interesting trade-off requires considering non-stationary relational contracts; we leave this for future work.

Recent theoretical work on relational contracts with subjective performance measures (Levin 2003, MacLeod 2003 and Fuchs 2007) has emphasized that the realized performance observed by the principal may be subject to noise, so that the principal and the agent have private information on what they observe. In this regard, a common theme is that in order to induce the principal to report perceived performance truthfully, the optimal contract must make the principal indifferent between reporting different performance levels. In relationships with a single agent, this tends to induce inefficiencies (the phenomenon of "money burning"). In this paper we have shown that excluding some agents is in fact optimal in most relevant cases. In particular, we have shown that with restricted competition the principal does not gain from punishing an agent that did not perform, because all agents are identical and replacing one with another brings no benefit. This means that if we had assumed a subjective performance assessment, we would have found analogous equilibria and results. The reason why no additional inefficient "money burning" is required – already pointed out by Fuchs (2007) – is that, unlike bilateral relationships, multilateral relationship with competing agents allow the principal to penalize the incumbent for poor performance without gaining anything itself: the gain goes instead to a competing agent. This maintains incentives for truthful performance reporting.

References

[1] Abreu, D., (1988), "On the Theory of Infinitely Repeated Games with Discounting", *Econometrica*, 56, 383-396.

[2] Affuso, L. and D. Newbery (2002), "The Impact of Structural and Contractual Arrangements on a Vertically Separated Railway", *The Economic and Social Review*, Vol. 33, 1: 83-92.

[3] Banfield, Eric, (1975), "Corruption as a Feature of Governmental Organization", *Journal of Law and Economics* 18, 587-605.

[4] Baker, G., Gibbons, R. and Murphy, K.J., (1994), "Subjective Performance Measures in Optimal Incentive Contracts", *Quarterly Journal of Economics*, 109, 1125-56.

[5] Bar-Isaac, H., (2005), "Imperfect Competition and Reputational Commitment," *Economics Letters*, 89(2), 167-173.

[6] Blume, A. and Heidhues, P. (2006), "Private monitoring in auctions," *Journal of Economic Theory*, 131(1), 179-211.

[7] Bull, C. (1987), "The Existence of Self-Enforcing Implicit Contracts", *Quarterly Journal of Economics*, 102(1), 147-159.

[8] Bulow, J. and Klemperer, P., (1996), "Auctions Versus Negotiations", *American Economic Review*, 86(1), 180-94.

[9] Europe Economics (2006), *Evaluation of Public Procurement Directives*, European Commission, DG Internal Market.

[10] Fuchs, W., (2007), "Contracting with Repeated Moral Hazard and Private Evaluations", *American Economic Review*, 97(4), 1432-1448.

[11] Fudenberg, D., D. K. Levine and E. Maskin, (1994), "The Folk Theorem in Repeated Games with Imperfect Public Information", *Econometrica*, 62, 997-1039

[12] Hart, O., (1995), *Firms, Contracts and Financial Structure,* Clarendon Press, Oxford.

[13] 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.

[14] Hörner, J. (2002), "Reputation and Competition", *American Economic Review*, 92(3), 644-663.

[15] Kelman, S., (1990), *Procurement and Public Management: The Fear of Discretion and the Quality of Government Performance,* AEI Press.

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

[17] Kranton, K., (2003), "Competition and the Incentive to Produce High Quality", *Economica*, 70, 385–404.

[18] Levin, J., (2002), "Multilateral Contracting and the Employment Relationship," *Quarterly Journal of Economics*, 117(3).

[19] Levin, J., (2003), "Relational Incentive Contracts", *American Economic Review*, 93(3), 835-847.

[20] Macaulay, S., (1963), "Non Contractual Relations in Business: A Preliminary Study", *American Sociological Review*, 28, 55-67.

[21] MacLeod, W. B. (2003), "Optimal Contracting with Subjective Evaluation", *American Economic Review*, March 2003, 92(1), 216-40.

[22] MacLeod, W.B. (2007), "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature*, Vol. 45(3), pages 595-628.

[23] MacLeod, W.B. and Malcomson, J.M., (1989), "Implicit Contracts, Incentive Compatibility, and the Involuntary Unemployment", *Econometrica*, 57(2), 447-480.

[24] MacLeod, W. B. and Malcomson, J.M., (1998), "Motivation and Markets", *American Economic Review*, 88(3), 388-411.

[25] Manelli, A. and Vincent, D.R., (1995), "Optimal Procurement Mechanisms", *Econometrica*, 63(3), 591-620.

[26] Rayo, L., (2007), "Relational Incentives and Moral Hazard in Teams," *Review of Economic Studies* 74(3), 2007.

[27] Rotemberg, J. and Saloner, G., (1986), "A Super-game Theoretic Model of Business Cycles and Price Wars during Booms", *American Economic Review,* 76(3), 390-407.

[28] Shapiro, C., (1983), "Premiums for High Quality Products as Returns on Reputation", *Quarterly Journal of Economics*, 98, 659-80.

[29] Shapiro, C. & Stiglitz, J. (1984), "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review* 74(3), 433–444.

[30] Skrzypacz A. and Hopenhayn H. (2004) "Tacit Collusion in Repeated Auctions," *Journal of Economic Theory* 114 (1), 153-169.

[31] Stiglitz, J. E., (1989), "Imperfect Information in the Product Market", in Richard Schmalensee and Robert Willig, eds., *Handbook of industrial organization*, 1. Amsterdam: Elsevier Science Publishers, 771–847.

[32] Strausz, R. "Planned Obsolescence and the Provision of Unobservable Quality," forthcoming the *Economic Journal*.

[33] Tirole, J., (1999). "Incomplete Contracts: Where Do We Stand?", *Econometrica*, 67(4), 741-782.

Appendix

Proof of Lemma 1. For any $n > 1$, substitute w from the binding agents' participation constraint $w + \beta(n)\pi(n) \geq 0$ into the buyer's objective function and, using $\beta(n) = \Pr[\theta_{it} = \theta'(n)] = 1/n$, we get $V_0 = v_0 \frac{1}{1-\delta} - \left\{E[\theta'(n)] + \theta_e(x) + k\right\} \frac{1}{1-\delta^x}$. The optimal C_0 simply follows from maximization of V_0.

The fee w is transferred before agent i learns its type θ_{it}; hence we are considering ex-ante participation for the agent. Allowing w to be paid after the agent learns its type would not alter our results qualitatively. In fact, in this case the principal may be able to extract some informational rents from an unique supplier, but still less than when allowing $n > 1$ agents to compete.

Furthermore, we are not considering extremely high values of k that would make the principal prefer not to screen at all and rather contract with a single agent, i.e. setting $n = 1$. This possibility would clearly make the analysis trivial both when $q = 0$ and when $q > 0$. ■

Proof of Lemma 2. Consider any auction taking place at t. Clearly, if any agent prefers to bid in t anticipating that it will deliver quality q if it wins, it will prefer to do so at any future auction in $t' > t$. Furthermore, we next show that in a SPA in which all agents plan not to cheat, truthful bidding is a (weakly) dominant strategy. Although in the proof we make use of the properties of the SPA and full information among agents, it will be clear that none of the results hinges on these assumptions.

Suppose agent i is the most efficient in t. When all the other agents bid truthfully and plan to deliver q, then by truthful bidding and planning not to cheat on q this agent i wins the auction and obtains the expected payoff indicated in the l.h.s. of the following

$$b_w - \theta_e(x) - \theta'(n) - c(q,x) + \{w + \beta(n)\pi(n)\} \frac{\delta^x}{1-\delta^x} \geq b_w - \theta_e(x) - \theta'(n) - B,$$

where the r.h.s. is instead the payoff if he does cheat. It is also immediate that i has no incentive to bid untruthfully and/or cheat on q. Hence, constraint (2) controls the incentives for the most efficient agent both ex-ante (at the bidding stage) and ex-post (at the quality-determination stage).

We consider now the possible deviation by agent i who is not the most efficient in t and may bid planning to cheat. If agent i anticipates delivering q in case it wins, he will bid $b_i = \theta_{it} + \theta_e(x) + c(q,x) - B$. On the contrary, if he plans to cheat his costs would be $\theta_{it} + \theta_e(x)$. If he bids planning not to cheat, he will not win in t, still obtaining a payoff equal to w plus

$\{w+\beta(n)\pi(n)\}\frac{\delta^x}{1-\delta^x}$, from future auctions. If he bids planning to cheat, let us consider the most favorable case to this agent, so that he bids the price of the most efficient agent (who was supposed to win in t) i.e. $b_w = \theta'(n) + \theta_e(x) + c(q,x) - B$ and he certainly wins. His payoff would be $w + b_w - \theta_{it} - \theta_e(x)$, so that he will prefer to bid not planning to cheat if

$$\theta_{it} - \theta'(n) + \{w+\beta(n)\pi(n)\}\frac{\delta^x}{1-\delta^x} \geq c(q,x) - B. \tag{10}$$

Since $\theta_{it} - \theta'(n) \geq 0$ it is immediate that (2) implies (10). ■

Proof of Lemma 3. Constraint (3) controls principal's incentives. If he is not satisfied the principal prefers retaining B and reverting to the outside option $\underline{V}_n$. Clearly, if $V_n = \underline{V}_n$ the principal faces no cost for reneging B and it must be $B = 0$. This is the case when $n < N$ since the principal can replace the current supplier with one from the $N-n$ pool of previously excluded agents. If instead $n = N$, then reneging B has the cost of losing one potential supplier so that $\underline{V}_N$ is the maximum payoff the principal can get with $N-1$ agents and then clearly $V_N \geq \underline{V}_N$. Notice that after a deviation from a contract C_n with $n = N$ agents, the possibility of turning to a contract that contemplates $n < N-1$ agents is dominated for the principal by a contract with $N-1$ agents. Indeed by so doing $(N-1)-n$ agents would be negatively affected by being excluded from the new relational contract and would not participate anymore in the future. Thus, the number of potential suppliers would be in fact n and not $N-1$, with no possibility of replacing a supplier in any case.

Furthermore, none of the $N-1$ agents would be negatively affected by the change to the relational contract by the principal after his own deviation and reducing the number of agents from N to $N-1$. Indeed, the optimal relational contract with open competition and $N-1$ would leave these agents with the same expected payoff.

As for participation constraint (4), starting with any $n > 1$ agents the principal has always the possibility of forgoing non-contractible quality, thus getting V_0^*. ■

Proof of Proposition 1. (i) Since $B = 0$ the principal maximizes

$$V_n = s(q)\frac{1}{1-\delta} - [E(\theta'(n)) + \theta_e(x) + k\mathcal{I}_{n>1}]\frac{1}{1-\delta^x} - \frac{n}{\delta^x}c(q,x)$$

w.r.t. q, n and x.

The optimal q_n^*, n^* and x_n^* satisfy the following conditions

$$\frac{ds(q)}{dq} = \frac{n}{\delta^x}(1-\delta)\frac{dc(q,x)}{dq}, \tag{11}$$

$$k\mathcal{I}_{n>1} \geq \{\theta_e - E[\theta'(n)]\} + \frac{n}{\delta^x}c(q,x)\ \ \frac{1-\delta^x}{\delta^x}, \tag{12}$$

$$\Delta\theta(n-1)\frac{1}{1-\delta^x} \geq c(q,x)/\delta^x \geq \Delta\theta(n)\frac{1}{1-\delta^x} \tag{13}$$

where $\Delta\theta(n) \equiv E[\theta'(n)] - E[\theta'(n+1)] \geq 0$ if $n > 1$ and $\Delta\theta(n) \equiv \Delta > 0$ if $n = 1$ with the scalar $\Delta > c(q,x)/\delta^x$ for any q and x.

Since the r.h.s. in (11) is positive and increasing in n and x, it follows that the optimal q_n^* implicitly defined by (11) is $q_n^* \leq q_{FB}$ and is decreasing in x and n.

If (12) is satisfied with a strict inequality then the optimal contract length x_n^* is infinite. Since, in the first best, the optimality condition for x would be the same but with last term in the r.h.s. (i.e. the distortion) equal to zero, it follows that $x_n^* \leq x_{FB}$. Notice also that if $n = 1$ then (12) is impossible and $x_n^* = 1$. Note that $x = \infty$ is now never optimal since this would imply that (12) is never satisfied and the optimal x would be $x = 1$, a contradiction, hence, $x_n^* \in [1, \infty)$. Furthermore, a larger n increases the r.h.s. in (12) thus inducing a smaller x and this is true not only for the curly bracket (as in the first best) but also due to the additional term in the r.h.s.

The optimal number of competing agents n^* is implicitly defined by the two inequalities in (13) where the l.h.s. is the cost increase of a unitary reduction of competing agents from n to $n-1$ and the r.h.s. is the cost reduction of a unitary increase of agents from n to $n+1$. The intermediate term is the effect on distortion in V_n induced by a unitary change (increase or reduction) in the number of agents in terms of quality. Increasing n^* by one more agent increases the distortion on quality with a negative effect on V_n that is larger than the efficiency gain on agents' cost. On the other hand, excluding one more agent from n^* generates an efficiency loss that has a more important effect on V_n than the reduction of distortion on quality. Condition (13) shows that n^* is larger (smaller) the smaller (larger) is the distortion $c(q,x)/\delta^x$ so that $n^* \leq n_{FB}$.

The relation between optimal x_n^* and n^* is finally characterized by even stronger substitutability than with first best since the terms induced by the distortion in the previous first order conditions all point in the same direction of stronger substitutability. From all these arguments it follows that since from (11) q_n^* is increasing in v, then n^* and x_n^* are both decreasing in v.

(ii) Since a larger B increases V_N and helps satisfy (3), the principal increases B as much as possible up to $c(q,x)$, i.e. $B = c(q,x)$, so that and the program becomes

$$\begin{array}{l} \max\limits_{q,x} \quad V_N \quad (= s(q)\frac{1}{1-\delta} - \left\{E[\theta'(N)] + \theta_e(x) + k\right\}\frac{1}{1-\delta^x}) \\ s.t. \\ (3): \; i.e. \; (V_N - \underline{V}_N)\delta \geq c(q,x). \end{array}$$

Since V_N is the same objective function in the first best, if q_{FB}, x_{FB} satisfy (3) then the optimal C_N^* is fully efficient. This possibility will be illustrated below. If instead with q_{FB}, x_{FB} (3) binds then the optimal C_N^* is distorted with (3) necessarily binding. In this case, let $\lambda \geq 0$ be the Kuhn-Tucker multiplier of constraint (3). The optimality conditions for q and x are respectively

$$\frac{\partial V_N}{\partial q} = \frac{\partial c(q,x)}{\partial q}\frac{\lambda}{\delta(1+\lambda)}, \quad \frac{\partial V_N}{\partial x} = \frac{\partial c(q,x)}{\partial x}\frac{\lambda}{\delta(1+\lambda)}.$$

Since $\partial c(q,x)\backslash\partial q \geq 0$, $\partial c(q,x)\backslash\partial x \geq 0$ it follows that $q_N^* \leq q_{FB}$, $x_N^* \leq x_{FB}$.

The derivation of the optimal C_N^* would require us to solve backward from the optimal contract C_1^* for $N = 1$ and increasing N. Although we will not follow this approach here, it is worth noticing that $V_N \geq \underline{V}_N$ and, if at the optimum (3) does not bind for a given N, then it also does not bind at the optimum for $N' \leq N$ since $V_N - \underline{V}_N$ is decreasing in N.

Notice that, as long as (3) does not bind, then V_N^* is increasing in N. Consider first a N such that (3) does not bind so that $C_N^* = C_{FB}$. This implies that constraint (3) can be written as

$$\left\{E[\theta'(N-1)] + \theta_e(x_{N-1}^*) + k\right\} \frac{1}{1-\delta^{x_{N-1}^*}} - \left\{E[\theta'(N)] + \theta_e(x_N^*) + k\right\} \frac{1}{1-\delta^{x_N^*}} > c(q_N^*, x_N^*)/\delta \tag{14}$$

From Lemma 1 we can have two cases: either $x_{N-1}^* = x_N^*$ or $x_{N-1}^* > x_N^*$. First, let $x_{N-1}^* = x_N^*$ so that the previous expression reduces to

$$\Delta\theta(N-1)\frac{\delta}{1-\delta^{x_N^*}} > c(q_N^*, x_N^*).$$

Since the l.h.s. decreases in N, there exists an $\underline{N} \in \Im_+$ so that for $N \geq \underline{N}$ the previous inequality is violated and the constraint (3) must bind and for $N < \underline{N}$ the inequality is satisfied so that $C_N^* = C^{FB}$. Clearly, the larger v, the higher q_N^* and the smaller $\underline{N}$. Second, consider now $x_{N-1}^* > x_N^*$ so that the first term in the l.h.s. of (14) is even larger than the second and the result holds *a fortiori*. Finally notice that since q_N^* is increasing in v, then $\underline{N}$ is a decreasing function in v.

The characterization of C_N^* for any $N \geq \underline{N}$ for which (3) binds goes as follows. When $N = \underline{N}$, then both $q_{\underline{N}}^*$ and $x_{\underline{N}}^*$ jump discretely below q_{FB} and x_{FB} respectively (i.e. the optimal quality and contract length with $N < \underline{N}$), as implied by the binding constraint (3). Furthermore, for $N > \underline{N}$, q_N^* and x_N^* must decrease in N. To see this, consider first $N = \underline{N} + 1$ and let $\tilde{V}_{\underline{N}}$ be $V_{\underline{N}}$ where we replace the $q_{\underline{N}}^*$ and $x_{\underline{N}}^*$ optimal for $\underline{N}$ with q_N and x_N. We then have

$$V_N - \tilde{V}_{\underline{N}} = \Delta\theta(\underline{N})\frac{1}{1-\delta^{x_N}} \geq V_N - \underline{V}_N$$

where the inequality comes from $\tilde{V}_{\underline{N}} \geq V_{\underline{N}}$ by definition. Using this with (3) we have, for $N = \underline{N}+1$,

$$\Delta\theta(\underline{N})\frac{1}{1-\delta^{x_N}} \geq V_N - \underline{V}_N \geq c(q_N, x_N)/\delta.$$

As argued above, since $\Delta\theta(.)$ is decreasing, q_N, x_N must be decreasing in N. This argument can be replicated for any pair $N(> \underline{N})$ and $N'(= N+1)$ thus implying the result on q_N, x_N, with $\lim_{N\to\infty} x_N^* = 1$ and $\lim_{N\to\infty} q_N^* = 0$. ■

Proof of Proposition 2. Clearly for $N \leq \underline{N}(v)$ the relational contract C_N^* is the overall optimum since it corresponds to the first best. From now on we consider $N > \underline{N}(v)$. The buyer has three options: either open competition getting V_N^*, restricted competition with $n > 1$ and a payoff V_n^* or setting $n = 1$ (i.e. negotiation) and a payoff V_1^*.

By the proof of Proposition 1 we have that if N is large then the distortion on open competition can be very large and, at the limit, it obliges the principal to set $q = 0$ since

$$\Delta\theta(N)\frac{1}{1-\delta^{x_N}} \geq V_N - \underline{V}_N \geq c(q_N, x_N)/\delta$$

and $\lim_{N\to\infty} \Delta\theta(N) = 0$. Hence, since for large N the payoff V_n^* with restricted competition does not depend on N, it follows that there exists a $\overline{N}(v)$ such that $V_n^* \geq V_N^*$ for $N \geq \overline{N}(v)$ with $\overline{N}(v)$ increasing in v since the larger v the higher $c(q,x)/\delta$.

Furthermore, when v is very small then principal's interest in quality tends to disappear so that he will prefer open competition; thus it must be that $\overline{N}(v) = \underline{N}(v)$ when quality is not an issue, i.e. $v = 0$, and the principal wants maximal competition and, furthermore, $\overline{N}(v) - \underline{N}(v)$ is increasing in v.

If instead N is intermediate, i.e. $\in (\underline{N}(v), \overline{N}(v))$, the comparison is ambiguous because both screening modes are characterized by distortions that cannot be compared in general. ■

Proof of Corollary 1. Suppose first the buyer cannot use B. Constraint (2) must still bind at the optimum and the only difference between open and restricted competition is the number of competitors: as in the proof of Proposition 1 the trade-off is that more competitors increase efficiency but also increase the distortion on quality.

Suppose now that the principal cannot use w. First notice that (1) is always satisfied. Hence, for any $n \leq N$ the principals's program consists in maximizing

$$V_n = [v(q) - \psi(q)]\frac{1}{1-\delta} - \left\{E[\theta''(n)] + \theta_e(x) + k\,\mathcal{I}_{n>1}\right\}\frac{1}{1-\delta^x}$$

subject to (2), i.e.

$$\beta(n)\pi(n) \geq \frac{1-\delta^x}{\delta^x}\left[c(q,x) - B\right]$$

and subject (3), where the cost component now depends on the cost of the second most efficient agent, i.e.$E[\theta''(n)]$ instead of $E[\theta'(n)]$.

With open competition, since a larger B has no effect on V_N and relaxes constraint (2), the principal optimally sets B so that (3) binds and (2) finally becomes

$$\beta(N)\pi(N) \geq \frac{1-\delta^x}{\delta^x}\left[c(q,x) - \delta\left(V_N - \underline{V}_N\right)\right].$$

In the case of restricted competition and negotiation, the reasoning is similar except for $B = 0$ so that the principal maximizes V_n subject to (2), i.e.

$$\beta(n)\pi(n) \geq \frac{1-\delta^x}{\delta^x}c(q,x).$$

If (2) does not bind at the optimum, open competition is preferred due to its lower costs, i.e. $E[\theta'(N)] < E[\theta''(n)]$ for any $n \in [1, N)$. However, if v is sufficiently high then (2) binds and implicitly defines the maximum enforceable quality $q(x,n)$. Comparing the two constraints with N and n agents, two effects are at play. First with more agents the l.h.s. is smaller since both $\beta(.)$ and $\pi(.)$ are decreasing functions. Second, the positive effect of B with open competition (which reduces the r.h.s. of constraint (2) by $(1-\delta^x)/\delta^{x-1}\ (V_N - \underline{V}_N)$) is decreasing in N. Hence, Proposition 1 applies. Furthermore, using w reduces the l.h.s. of (2) and makes the desirable effects of reducing the number of competing agents on q even stronger.

Finally, if the principal cannot use either B or w, our previous analysis combines and the results holds. ■

Proof of Lemma 4. (i) Suppose first the principal cannot use w or B and consider a relational contract so that the maximum enforceable quality when agents do no collude is $q(n,x)$ implicitly defined by (7). A necessary condition for collusion to be incentive-compatible is $\tilde{\pi}(n) \geq \pi^*(n)$. Hence, from (7) the result immediately follows. Assume now that the principal can use B but not w. Collusion is stable if

$$[B + \beta(n)\tilde{\pi}(n)]\,\frac{\delta^x}{1-\delta^x} \geq D + \max\{B + \beta(n)\pi^*(n)\frac{\delta^{x^*}}{1-\delta^{x^*}}, c(q,x)\}$$

and the current supplier does not cheat on quality if

$$[B + \beta(N)\tilde{\pi}(N)]\,\frac{\delta^x}{1-\delta^x} \geq c(q,x) - B.$$

Comparing competition and collusion in the case of open competition, we notice that a unitary increase of B with collusion relaxes the constraint on quality by $1 + \frac{\delta^x}{1-\delta^x}$ which is more than would happen with a similar increase of B with competition. Hence, ceteris paribus the principal can enforce higher quality with collusion.

If the principal can now use w but not B, the cartel's stability constraint is

$$[w + \beta(n)\tilde{\pi}(n)]\,\frac{\delta^x}{1-\delta^x} \geq D + \max\{[w^* + \beta(n)\pi^*(n)]\,\frac{\delta^{x^*}}{1-\delta^{x^*}}, c(q,x)\},$$

which substituting $w^* = c(q^*,x^*)\frac{1-\delta^{x^*}}{\delta^{x^*}} - \beta(n)\pi^*(n)$ becomes

$$[w + \beta(n)\tilde{\pi}(n)]\,\frac{\delta^x}{1-\delta^x} \geq D + \max\{c(q^*,x^*), c(q,x)\}.$$

Hence, since $D+\max\{c(q^*,x^*), c(q,x)\} \geq c(q,x)$, the incentive-compatibility constraint for collusion implies the supplier's incentive-compatibility constraint for quality, i.e. $[w + \beta(n)\tilde{\pi}(n)]\,\frac{\delta^x}{1-\delta^x} \geq c(q,x)$ (which is the same whether agents collude or not).

(ii) Assume that the principal can use both B and w. The incentive-compatibility constraint of collusion (8) then becomes

$$[w + B + \beta(N)\tilde{\pi}(N)]\,\frac{\delta^x}{1-\delta^x} \geq D + \max\{B, c(q,x)\}. \tag{15}$$

For our argument it suffices to consider here $\delta^x \geq 1/2$ so that a larger B actually strengthens collusion. Constraints (15) and (9) can be combined, producing the single constraint,

$$[w + B + \beta(N)\tilde{\pi}(N)] \frac{\delta^x}{1-\delta^x} \geq \max\{D + \max\{B, c(q,x)\}, c(q,x) - B\}.$$

Suppose $c(q,x) - B \geq D + max\{B, c(q,x)\}$ which requires $c(q,x) \geq B$ because $D + max\{B, c(q,x)\} > 0$, so that the above becomes $-B \geq D$: this is impossible since $D > 0$. Hence, it must be that $c(q,x) - B \leq D + max\{B, c(q,x)\}$ and the relevant constraint is (15). This immediately shows that the constraint on quality is not a concern once collusion is in place because it is itself implied by the stability of collusion. Hence, higher quality can be enforced with collusion than with competition. ■

Proof of Proposition 4. (i) Assume the principal cannot use w. First notice that (8) can be written as

$$[B + \beta(n)\tilde{\pi}(n)] \frac{\delta^x}{1-\delta^x} \geq D + \max\{B + \beta(n)\pi^*(n)\frac{\delta^{x^*}}{1-\delta^{x^*}} - [c(q,x) - B], B\} + c(q,x) - B$$

which implies the constraint on quality (9), i.e.

$$[B + \beta(n)\tilde{\pi}(n)] \frac{\delta^x}{1-\delta^x} \geq c(q,x) - B.$$

Hence, if v is large enough, the principal wants to enforce a higher q and then constraint (8) binds. As previously shown this, together with $\tilde{\pi}(n) \geq \pi^*(n)$ immediately implies that, for any B, the principal prefers inducing collusion so as to have a larger q. In addition, since with $n = N$ and colluding agents B is part of agents' profits, then for any $B > 0$ the previous result holds a fortiori.

Consider now the possibility that $n = 1$. In this case, the principal induces the largest attainable expected rent to the (single) supplier and then creates the strongest incentives for quality q. Indeed, consider the rent of the single agent with $n = 1$ and let us compare it with the expected rent with $n > 1$ colluding agents. With probability $1/n$ the single agent with negotiation will have the same efficiency as the most efficient agent in the cartel of n agents, thus earning the same profit. With complementary probability it will not be as efficient but still earn a profit contrary to what would happen for that agent being in the cartel.

As a last point, imagine now that the principal also cares about efficiency and therefore maximizes

$$V_n - \gamma E[\theta_{it} + \theta(x) + c(q,x)],$$

where $\gamma \geq 0$ is some weight to the expected cost of competition. Consider then the case in which the value attached to quality is great so that the principal either procure with colluding agents or with negotiation. Since for given q and x the cost with negotiation is

$$E[\theta_e + \theta(x) + c(q,x)] \geq E[\theta'(n) + \theta(x) + c(q,x)]$$

where the r.h.s. is the cost with colluding agents, then for sufficiently large γ the principal prefers colluding agents.

(ii) Finally, assume now that the principal can use the fixed fee w. Following the same steps as above, the relevant constraint with collusion is the cartel's stability so that

$$w = [D + \max\{B, c(q, x)\}] \frac{1-\delta^x}{\delta^x} - B - \beta(n)\tilde{\pi}(n)$$

(recall that with competition $c(q^*, x^*) = B^*$). This transfer is certainly larger than that with competing agents if the principal must set $B = 0$ (if B is not practicable or if $n < N$) since with competition for any $B \geq 0$ we have

$$w = \max\{c(q, x) - B, 0\} \frac{1-\delta^x}{\delta^x} - \beta(n)\tilde{\pi}(n).$$

Hence, if the principal cannot use B, then collusion is immediately dominated by competition.

If $B > 0$ then the comparison of the two previous transfers shows a positive effect of B with collusion that is absent with competition: with colluding agents the principal actually pays B which allows the reduction of w to all N suppliers with a net benefit in V_N for the principal equal to $\frac{1}{1-\delta^x} B(N-1)$:

$$\begin{aligned} V_N &= [v(q) - \psi(q)] \frac{1}{1-\delta} - \left\{E[\theta'(N)] + \theta(x) + k\right\} \frac{1}{1-\delta^x} + \\ &\quad -N \frac{1}{1-\delta^x} D - \max\{B, c(q, x)\} \frac{N}{\delta^x} + \frac{1}{1-\delta^x} B(N-1). \end{aligned}$$

If the derivative of V_N w.r.t. B is negative then $B = 0$ and again collusion is dominated, as previously shown. Suppose $B \leq c(q, x)$ so that the derivative of V_N is positive. The benefit of using $B > 0$, i.e. $\frac{1}{1-\delta^x} B(N-1)$, has to be compared with the cost of collusion, i.e. the increase of w by $D\frac{1-\delta^x}{\delta^x}$ with an associated cost in V_N equal to $-N\frac{1}{1-\delta^x}D$. If N is sufficiently large, then following the same steps as in the proof of Proposition 1 part (ii), B is small and then $\frac{1}{1-\delta^x} B(N-1) \leq N\frac{1}{1-\delta^x}D$: the cost prevails. If instead N is small, then inducing collusion is optimal if $(N-1)/N > D/B$. ■