Reciprocity and the Costs of Authority Relationships

Nabil I. Al-Najjar*

and

Chris Forman†

May 1999

Revised: November 1999

Abstract

Authority relationships are viewed as reciprocal exchange in which a principal offers rents in return for subordinates' compliance with his authority. These rents induce compliance by creating a collective action problem among subordinates so they free-ride on each other in challenging the principal's authority. As a consequence of the payment of these rents, the cost of exercising authority may distort the principal's ex ante choice of internal authority relationship as an organizational form, relative to market exchange or formal, arms-length contracting.

*Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. <al-najjar@nwu.edu>

†Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. <c-forman@nwu.edu>

1 Introduction

The distinctive feature of firms and organizations is their reliance on authority to replace markets and formal, arms-length contracting as methods of internal resource allocation. Interactions within firms typically consist of bosses, employers and organizational leaders telling subordinates what to do, with .their directives carried out without active or overt resistance.

The pervasiveness and significance of authority is hardly in doubt: Coase (1937), Simon (1951), and Arrow (1974) are classic works identifying firms and organizations as authority structures. The central role of authority relationships is also shared by modern economic theories of organizations[1] as well as legal and sociological literatures on employment relationships.[2]

The advantages of authority are often obvious: authority can dramatically reduce the coordination, bargaining and renegotiation costs. What is more puzzling is why the exercise of authority entails any costs at all. Why aren't all contractual relationships organized based on one party's direct authority and control over others?

This paper explores the *reciprocal* nature of authority as an answer to this question. We present a model in which subordinates offer compliance in exchange for rents and boundaries on the scope of activities over which authority is exercised. Our main conclusion is that this give-and-take represents a major cost of exercising authority. In particular, it creates a trade-off between a potentially more efficient, but more costly, authority relationship with employees and arms-length relationships through contracts and mar-

[1] This is a central theme of the transaction cost literature, following Williamson (1985), the property rights approach of Grossman and Hart (1986), and the firms as a carrier of reputation, as in Kreps (1990).

[2] In the sociological literature, see Halaby (1986) who elaborates on the central role played by workplace authority.

kets.

We consider a relationship between a principal and N subordinates characterized by private information and irreversibilities (lock-in). The principal sets the terms of the relationship in the form of a contract specifying compensation, the directives subordinates are expected to follow, the scope of authority, etc. A key contractual imperfection is the parties' limited contractual commitment: once locked into the relationship, subordinates have an incentive to challenge the principal to extract better terms than initially agreed to, and the principal is tempted to capitulate when facing a collective challenge.

Subordinates may collude against the principal to extort him for better terms. We allow rich collusion possibilities in the form of direct mechanisms where subordinates reveal their types and exchange transfers. The principal *exercises authority* when subordinates voluntarily refrain from such challenges and comply with the terms of the relationship even though no formal mechanisms exist to enforce these terms.

Our analysis builds on the tension between subordinates' collective benefit from a successful challenge of the principal's authority and their incentive to free-ride on each other in such challenges. Reciprocal exchange as basis of authority arises from the peculiar nature of this free-rider problem: unlike standard public good settings, where the costs and benefits of contributing are exogenously given parameters, the cost of challenging authority here is endogenously controlled by the principal. In fact, the principal largely *creates* the collective action problem by using rents to manipulate subordinates' incentives. As an extreme example, consider a principal who leaves subordinates indifferent between challenging and complying. Since subordinates have little to risk losing, this is like a public good problem in which the cost of contributing to its provision is zero. No free-riding has to occur; indeed, it is easy to devise mechanisms to induce a collective action that destabilizes authority.

In our model, a successful exercise of authority requires the principal to, in effect, bribe subordinates with rents that increase the cost of challenging, to which subordinates reciprocate by voluntarily complying with his authority. We provide a simple model explaining how the choice of authority structure, magnitude of rents, boundaries on authority, and so on, are endogenously determined by primitives like private information, lock-in, degree of commitment, and available means of collusion and coordination.

Two features of our model are worth emphasizing. First, we emphasize the role of *robust incentives*. The rich set of subordinates' collusive interactions represents, to the principal, considerable strategic uncertainty over which he has little control. On the other hand, a stable authority suggests a relationship where subordinates see no viable alternative to compliance, that compliance appears as the 'obvious thing to do.' To capture this, we assume that the principal hedges against this strategic uncertainty by designing authority structures robust to subordinates' *optimal* collusive mechanism. In other words, our robustness requirement is that the principal designs his authority structure to hedge against the worst case scenario compatible with subordinates' self-interest. This hedging leads the principal to pay rents even though challenges rarely occur. In the model, these rents represent the cost of maintaining a stable authority relationship.

Second, the mechanism-design approach to collusion enables us to examine alternative assumptions about the collusive mechanisms available to subordinates. For example, coercive measures (e.g. unions, powerful social norms and institutions), subsidies (e.g. a union war chest), or a divide-and-conquer strategy to weaken subordinates can be conveniently introduced as constraints on the set of allowable mechanisms. We build on the insights of the literature on public-good provision and free-riding under asymmetric information, in particular the works of Rob (1989), Mailath and Postlewaite (1990) and Ledyard and Palfrey (1994), to examine how the authority relationship is affected by changes in primitives of the environment. In particular, we use the model to explore how the cost of exercising authority may distort the principal's choice between an internal authority relationship vs. alternative organizational forms, and how the outcome of the relationship is affected by such things as the presence of collusive organizations (such as unions), external subsidies, the cost of sanctioning subordinates, and the extent of parties' lock-in due to, say, relationship-specific irreversible investments. Section 5 discusses related literature.

2 The Model

A principal enters into a relationship with N subordinates , each generating a surplus $S > 0$. Interaction occurs in two stages: in the ex ante *contracting stage* the principal chooses between an arms-length interaction and an authority relationship; in the latter case, he offers a compensation scheme. If subordinates accept, play proceeds to a *collusion game* in which subordinates may collude to extort the principal for better terms.

2.1 Surplus, Actions and commitment

The principal offers a compensation scheme (w, w') , common to all subordinates, where w represents an upfront reward to be paid immediately, while w' is a sanction imposed if a subordinate subsequently fails to comply.

If the scheme is accepted, parties enter into a relationship where each subordinate takes an action a_n , observed by the principal, indicating either compliance, a^*, or challenge, a^0. This interpretation of actions will be borne by their effects on payoffs. Compliance may be a contingent actions (i.e. of the form 'take action a_x when contingency x arises') indicating the principal's discretion to direct the subordinate to take a more elementary actions as contingencies arise (in the sense of Simon (1951)).

The key contractual imperfection driving our analysis is the principal's inability to fully commit to stick by the terms initially offered, (w, w'). One motivation for this assumption is that ex ante the principal commits to w, but for flexibility reasons he does not completely tie down his hands, leaving room for undoing this commitment should it be warranted by ex post contingencies. To model this, we assume a *stochastic commitment* technology: with probability ρ the principal upholds his commitment regardless of what subordinates do, and with probability $1-\rho$ the principal's commitment may be undone.[3]

2.2 Reservation values and lock-in

Each subordinate has an *ex ante* reservation value $\bar{U} \geq 0$, while the principal's per capita reservation value is $\bar{V} \geq 0$. Once parties agree to enter the relationship, they are locked-in, a fact reflected by new *ex post* reservation values, which we assume to be 0.

The values of $\bar{V}$ and $\bar{U}$ thus measure the strength of the lock-in effects. The importance of these effects in organizational and competitive contexts has been recognized, most prominently in the work of Williamson (1985). For example, the principal may have ex ante the option of procuring services through competitive markets or arms-length contracting, but this option is lost or severely diminished once committed to the relationship with the

[3]Here ρ is exogenous, as we do not model explicitly the trade-off between commitment and flexibility. Our comparative static analysis in Section 4 suggests how ρ should be set, had the principal had the freedom to do so.

subordinates. Subordinates may also forfeit outside opportunities through sunk, relationship-specific investments.

2.3 Sanctions

Since subordinates' actions are perfectly observed, it follows (under our other assumptions on payoffs introduced later) that the principal should optimally set w' equal to the harshest possible punishment consistent with subordinates ex post reservation value of 0. Thus, the worst sanction the principal can impose is taking back the reward w, which one may interpret as firing the subordinate, leaving him with the possibly deteriorated ex post outside options. To simplify the exposition, we assume at the outset that $w' = -w$, so the only relevant component of the incentive scheme is w.

Sanctioning challenges is also costly to the principal, due to (unmodeled) effects on the disruption of production, adverse effects on other workers' morale, and so on. Such costs have been emphasized by Akerlof (1982), and are consistent with stylized facts about firms' reluctance to implement disciplinary measures or fire employees (Bewley (1998) reports on this extensively).

In general, the cost of sanctioning challengers may depend, among other things, on the number of challengers and the total number of subordinates. Here we assume, for simplicity, that this cost is linear, taking the form gK, where K is the number of challengers and $g > S$ is the cost of meeting a each challenge, so meeting a challenge dissipates all the value to the relationships, and is therefore never efficient.

2.4 The public-good nature of authority

Given the principal's lack of full commitment, subordinates' behavior is determined by two opposing incentives: an incentive to collude to challenge the principal's authority, and an offsetting incentive to free-ride on other subordinates' challenges.

This free-rider problem hinges on the public-good nature of a successful challenge: by capitulating with one subordinate, the principal undermines his authority in dealing with others. We make the (extreme) simplifying assumption that if the principal capitulates one subordinate, then he must capitulate with all others. In particular, it is not possible to exclude a subordinate from the benefits of a successful challenge, even though he might have dodged the costs of participating in that challenge.[4]

Formally, in the collusion stage the principal faces a profile $(a_1, \ldots, a_N)$ of compliance/challenge, against which he either: (1) capitulates, forfeiting the surplus S, an action denoted $\gamma = 1$; or (2) upholds the terms of the relationship by sanctioning those who challenge him, $\gamma = 0$. The stochastic commitment discussed above implies that $\gamma = 0$ with probability at least ρ.

If the principal's commitment can be undone (an event that has probability $1 - \rho$), then sanctioning challengers yields a payoff (net of w which is sunk at this stage)

$$NS - Kg,$$

while if he capitulates, he forfeits the surplus S and nets 0. The principal's decision is thus determined by the least number of challengers $\bar{K}$ such that $\bar{K} \geq \frac{NS}{g}$: he capitulates if there are $K \geq \bar{K}$ challenges, and defends his

[4]The logic of the argument would hold if, say, challengers appropriate a portion of the benefits, provided there remains a non-vanishing residual public good component.

authority otherwise.

2.5 The ex ante contractual choice and the commitment benchmark

It is useful to identify the *ex ante efficient* outcome in our setup. The surplus generated by entering into the relationship is S, while the combined value of the parties' outside options, $\bar{U}+\bar{V}$. If $S \leq \bar{U}+\bar{V}$, parties are better off pursuing their outside options. For example, the principal may be able to more efficiently procure services through arms-length contracting or market exchange instead of direct interaction with subordinates. The problem is interesting when

$$S > \bar{U}+\bar{V}, \tag{$*$}$$

in which case the principal faces a genuine trade-off between:

- a more efficient internal authority relationship, but one in which maintaining authority entails costs to keep subordinates incentives to challenge under control; and
- less efficient outside opportunities which do not entail such costs.

Clearly, the principal's decision hinges on the cost of authority, an issue we analyze in detail in the next section.

In studying the cost of authority, we find it useful to refer to the full commitment benchmark. Write $e \in \{0,1\}$ for the principal's decision to opt out and earn the outside option $\bar{V}$, or enter an authority relationship with subordinates. If the principal could commit to meet every challenge, then no challenges ever occur, and the principal's problem (on a per capita basis) is:

$$\textbf{Program } (P_0)\textbf{:} \quad \max_{e,w} e\,(S-w)+(1-e)\bar{V}$$

subject to the ex ante individual rationality constraints:

$$S - w \geq \bar{V} \quad (1)$$

$$w \geq \bar{U}. \quad (2)$$

Full commitment ensures that there is no cost to the exercise of authority. Our assumption of perfect observability of actions and that authority is efficient ($S > \bar{U} + \bar{V}$), imply that the principal sets $e = 1$, holds subordinates to their reservation value $w = \bar{U}$, and collects the entire ex ante surplus $S > \bar{U} + \bar{V}$. This outcome is ex-ante efficient.

3 Commitment and collusion

In this section we model collusion among subordinates. Rather than introduce the complex details of the collusion game explicitly, we use instead direct mechanisms in which subordinates reveal their private information and exchange transfers. Throughout, we assume that a specific w is given and sunk. In Section 4 we examine how collusion influences the ex ante setting of w.

3.1 Subordinates' types and payoffs

To introduce free-riding among subordinates, we follow the public good literature (Rob (1989), Mailath and Postlewaite (1990) and Ledyard and Palfrey (1994)) by assuming that subordinates have private information about some relevant characteristics. Specifically, we assume that subordinate n's valuation of the surplus takes the form $t_n S$, where $t_n \in T_n = \{0, 1\}$ is the subordinate's privately known type. Types are i.i.d., with type $t_n = 1$

having probability, $1 > p > 0$.[5]

If the principal defends the scheme w, i.e. $\gamma = 0$, subordinate n receives a wage adjustment $w(a_n)$:

$$w(a_n) = \begin{cases} 0, & a_n = a^* \\ -w, & a_n = a^0. \end{cases}$$

If the principal capitulates ($\gamma = 1$), subordinate n receives $t_n S$. Agents are risk neutral, so their expected payoff in the collusive mechanism is:

$$U(a_n, t_n, \gamma, w, c_n) = (1-\gamma)w(a_n) + \gamma t_n S + c_n,$$

where c_n represents transfers used to sustain collusion .[6]

3.2 Collusive Mechanisms

Let T and A denote the sets of possible type and action profiles, with generic elements t and a, respectively. A mechanism (σ, c) is a pair of functions $\sigma : T \to A$ and $c : T \to \Re^n$, where the nth components, $\sigma_n(t)$ and $c_n(t)$, denote subordinate n's action and transfer, respectively. With some abuse of notation, define subordinate n's utility under such a mechanism when he truthfully reports his type:

$$U(t, \sigma, c) = U(\sigma_n(t), t_n, \gamma(t), w, c_n(t))$$

We also define $U(t_n^{'}; t, \sigma, c)$ to be the utility when he mis-reports his type to $t_n^{'}$.[7] Consider the following benchmark requirements:

[5] Throughout the paper, we assume parties to be risk neutral.

[6] Note that we are assuming that the agent does not derive any direct benefit from the action he takes; his action affects his payoff only in so far as it influences the principal's behavior towards him. Section 4 examines the implications of dropping this assumption.

[7] That is, $U(t_n^{'} : t, \sigma, c) = U(\sigma_n(t_{-n}, t_n^{'}), t_n, \gamma(t_{-n}, t_n^{'}), w(\sigma_n(t_{-n}, t_n^{'})), c_n(t_{-n}, t_n^{'}))$.

Assumption 1: (σ, c) satisfies

1. *Budget Balance:*

$$\text{for every type profile} \quad t, \sum_n c_n(t) \leq 0, \tag{BB}$$

2. *Individual rationality:* For every n, t_n

$$E[U(t, \sigma, c)|t_n] \geq E[\gamma t_n S|t_n]. \tag{IR}$$

3. *Incentive compatibility:* For every n, t_n,

$$E[U(t_n^{'}; t, \sigma, c)|t_n] \leq E[U(t, \sigma, c)|t_n] \tag{IC}$$

Discussion: These are standard conditions in the public goods literature, including Rob (1989), Mailath and Postlewaite (1990a), and Ledyard and Palfrey (1994). In our context, the mechanism collects contributions and pays subsidies to finance the collective challenge to produce the 'public good' γ. Budget balance (BB) simply says that no third party subsidizes subordinates in their challenge of the principal. In contrast to standard public good problems, the cost of contribution in our model is the magnitude of the sanction, $-w$, which is endogenously set by the principal at an earlier stage. The 'technology of provision' $\gamma(t)$ reflects the principal's strategic decision, rather than an exogenously given function.

3.3 Robust incentives

To understand how the principal sets w ex ante, we need to make assumptions about his expectations about the collusive mechanism adopted by subordinates. One such mechanism is 'do nothing:' all subordinates comply and $c_n(t) = 0$ for all n and t. Although this satisfies A.1, we argue that for the principal to expect this mechanism to prevail in the collusion stage is unduly optimistic. Given that the principal has little or no control over how subordinates interact ex post, it is reasonable to expect him to hedge against a broad range of collusive mechanisms. We will in fact assume that the principal structures the relationship based on the expectation that subordinates will exhaust all feasible, incentive compatible collusive mechanisms in which they voluntarily participate.

Definition 1 *A mechanism* (σ', c') improves (strictly) *on* (σ, c) *if for every* t, $\sum_n E[U_n(t, \sigma', c')] \geq \sum_n E[U_n(t, \sigma, c)]$ *(>). A mechanism* (σ, c) *satisfying A.1 is* efficient *if there is no mechanism* (σ', c') *satisfying A.1 that strictly improves on it.*

We say the principal provides subordinates *robust incentives* if he sets w based on the expectation that subordinates will use an efficient collusive mechanism. Thus, although the principal hedges against a worst case scenario, his pessimism is realistic in the sense that he takes into account the factors impeding subordinates' collusion by assuming that collusion is organized subject to A.1.

3.4 Characterizing Efficient Collusive Mechanisms

We first show that in searching for efficient collusive mechanisms, we may restrict attention to a very simple class of mechanisms. First we introduce the following notation: let $K(t)$ be the (possibly random) number of challenges at the type profile t. A profile of actions is an *effective challenge* if $\bar{K}$ or more subordinates challenge in that profile, i.e. if the profile causes the principal to capitulate. Let $h(t)$ denote the number of type 1 subordinates in the profile t and define $q(h) = P\{K(t) \geq \bar{K} | h = h(t)\}$ to be the probability that an effective challenge is launched given that the type profile is one with h type 1 subordinates.

Definition 2 *A mechanism* (σ, c) *is* simple *if for every type profile* t*: (1)* $K(t) \in \{0, \bar{K}\}$*; and (2) there is* $\bar{h}$ *such that* $q(h) = 1$ *if* $h > \bar{h}$*,* $q(h) = 0$ *if* $h < \bar{h}$*, and* $q(\bar{h}) \in [0, 1)$*.*

The idea of simple mechanisms is motivated by Ledyard and Palfrey (1994). Roughly, such mechanisms launch only effective challenges, and they do so according to an $\bar{h}$-majority rule ($q(\bar{h})$ is unrestricted).

The following proposition characterizes efficient mechanisms: Let $d(h) = \frac{\bar{K}\rho w}{h}$, the per capita cost of challenge in a simple mechanism in which the cost of challenging is funded solely by type 1's, and let δ_h denote the probability of h successes in $N-1$ trials where the probability of success in each trial is p.

Proposition 1 *Every efficient mechanism satisfying A.1 is simple, and sets* $\bar{h}$ *to be the smallest integer satisfying*

$$\delta_{\bar{h}} \left[(1-\rho)S - d(\bar{h}+1) \right] \geq \sum_{h=\bar{h}+1}^{N-1} \delta_h \; d(h+1). \tag{3}$$

To get the intuition of how an optimal mechanism works, assume that this mechanism treats subordinates of the same type symmetrically, distributes the cost of challenges equally among type 1 subordinates for each h, and sets $q(\bar{h}) = 0$.[8] For $h > \bar{h}$, each challenger incurs an expected loss of ρw, so the total burden of a collective challenge is $\bar{K}\rho w$. Each subordinate of type 1, on the other hand, benefits by $(1-\rho)S$, while his share of the cost is $d(h)$.

[8] The Appendix shows that the first two assumptions are without loss of generality. The assumption $q(\bar{h}) = 0$ is inconsequential since we are interested in the cutoff $\bar{h}$, rather than the value of q at it.

Had there been no private information and free riding, efficiency would have required that $\bar{h}$ be set to the smallest integer satisfying:

$$(1-\rho)S \geq d(\bar{h}+1). \tag{E}$$

The difference between equations 3 and (E) captures the essence of the free-rider problem: Equation (E) requires challenges to be launched as long as their collective benefit outweigh their cost. By contrast, LHS of equation 3 is the expected gain when the launch of an effective challenge hinges on this subordinate truthfully reporting his type, an event that occurs only when there are exactly $\bar{h}$ type 1 reports by the remaining $N-1$ subordinates. This is offset by the RHS, which represents the cost of truthtelling: when $h \geq \bar{h}+1$, this subordinate must help pay for the cost of a challenge that would have been launched regardless of his report. The RHS of equation 3 thus represents his incentive to free-ride on others' collective challenge.

4 Rents and the costs of authority relationship

This section examines implications of subordinates' collusion on the principal's ex ante choice between arms-length contracting and authority and (in the latter case) the compensation he offers subordinates.

4.1 Optimal choice of authority structure

A principal who expects subordinates to use an efficient collusive mechanism faces an unconditional probability of challenge:

$$Q(w; N, S, \rho, g, p) = q(\bar{h})P(\bar{h}) + \sum_{h>\bar{h}} P(h),$$

where $P(h)$ is the probability of h successes in N trials. Since no challenge is launched unless it is effective, the principal's ex ante expected payoff is, conditional on entering the relationship, is:

$$V(w) = S - w - Q\left[(1-\rho)S + \rho g\frac{\bar{K}}{N}\right]. \tag{4}$$

The principal's ex ante choice between entering an authority relationship and exercising his outside option is determined by the following constrained maximization problem:

$$\textbf{Program (P):} \ \max_{e,w} \ eV(w) + (1-e)\bar{V}$$

subject to the constraints:

$$eV(w) + (1-e)\bar{V} \geq \bar{V} \tag{5}$$

$$w \geq \bar{U}. \tag{6}$$

$$Q = Q(w). \tag{7}$$

Constraints 5 and 6 are virtually identical to the ex ante participation constraints introduced earlier (Program P_0, p. 9). They reflect the ex ante opportunities forfeited by entering into the relationship. New is the constraint (7) that captures subordinates' collusion. Under this constraint, reducing w no longer leads to an unambiguous increase in the principal's payoff, since doing so increases the probability that subordinates challenge through the function $Q(w)$ in constraint (7). The principal may then be willing to offer better terms than required by subordinates participation in order to lower the probability that subordinates challenge him later.

Proposition 1 provides a closed-form expression for $Q(w)$, explaining how its value depends on the primitives of the environment and with the principal's choice of w. Figure 1 displays a typical shape of $Q(w)$, which is downward sloping in w, and concave on an interval $[0, w]$, $w \leq S$. The principal's payoffs determine a linear, downward sloping isoprofit lines with values increasing towards the origin. The optimal $w - Q$ combination is determined by the tangency of these two curves.

The effects of changes in the primitives of the environment can be seen directly through examination of (7). For example, for every h, δ_h decreases as N increases, lowering the LHS of (7) and requiring $\bar{h}$ to compensate. Since higher $\bar{h}$ lowers $Q(w)$ for every w, the challenge constraint of Figure 1 shifts downward and toward the origin, indicating that the principal's payoff unambiguously increases. An increase in the commitment technology ρ has a similar effect, lowering the LHS of (7) while decreasing the RHS, again lowering the probability of challenge $Q(w)$ for every w.

4.2 Rents

An important consequence of our analysis concerns the payment of rents:

Proposition 2 *If $\bar{V} > 0$ and (e, w) solves Program (P), then $e = 1$ implies $w > 0$.*

Proof: If $w = 0$, then sanctions are irrelevant for the subordinates' decision problem. In particular, the mechanism: $a_n = a^0$ and $c_n = 0$ for every n and t satisfies A.1, and yields $Q = 1$. But in this case $V(w) = S - w - Q[(1-\rho)S + \rho g \frac{\bar{K}}{N}] = S - [(1-\rho)S + \rho g \frac{\bar{K}}{N}]$ which, by the definition of $\bar{K}$, is non-positive. Since $\bar{V} > 0$, the constraint 5 binds hence $e = 0$. ■

With $w = 0$, subordinates face a problem of providing a public good (challenging) where the cost of contributing is zero. Since they have nothing to lose by challenging, one can easily find mechanisms that generate an effective challenge with probability 1. Thus, if the principal enters in an authority relationship, the w offered must be high enough to create a free-rider problem among subordinates.

The proposition has the following implication on whether subordinates are offered *ex ante rents*

$$r = w - \bar{U}$$

in excess of their ex ante reservation value. It is convenient to consider the optimal wage $\hat{w}$ of Program P under the additional assumptions that there is no subordinates' lock-in and that the principal is forced to enter into an authority relationship, (that is, $\bar{U} = 0$ and $e = 1$). Proposition 2 implies that $w \geq \hat{w} > 0$, while subordinates' participation forces $w \geq \bar{U}$, so $w = \min\{\bar{U}, \hat{w}\}$. Ex ante rents thus depend on the relationship between subordinates' lock-in, $\bar{U}$, relative to the $\hat{w}$ needed to generate the optimal level of free-riding among subordinates:

1. *Canceling hold-up problems*: If $\bar{U} > \hat{w}$, then subordinates participation forces $w = \bar{U}$, and no rents are paid. This occurs when there is substantial subordinates' lock-in, so subordinates are held-up by the principal because of the potential loss of their ex ante outside option upon being locked into the relationship. This hold-up problem 'cancels' the original hold-up of the principal by subordinates, thus no rents are paid.

2. *Rents for compliance*: If $\bar{U} < \hat{w}$, then subordinates' participation constraint (6) does not bind, and w is set so that $w = \hat{w} > \bar{U}$. In this case the principal pays strictly positive rents to raise subordinates' cost of challenging. In 'exchange' for these rents subordinates offer compliance in the sense of challenging less frequently than would have been in their collective interest.

For concreteness, we consider the following:

Numerical example (1): Fix the model's parameters with $N = 100$ subordinates, a surplus of $S = 1$, principal's commitment $\rho = 0.5$, and probability of type 1 subordinates $p = 0.5$. Assume a cost of sanctioning $g = 2$, so $\bar{K} = 50$. The principal's challenge constraint 7 and a typical isoprofit line are depicted in Figure 1. We numerically solved Program P under the assumptions $\bar{U} = 0$ and $e = 1$ to find a wage of $\hat{w} = 0.29$, a critical threshold of challengers $\bar{h} = 55$ set by the optimal simple mechanism, and an unconditional probability of challenge of $Q = 0.14$. The per capita ex ante payoff of the principal from entering an authority relationship is 0.57.

To illustrate the effects of ex ante participation constraints, for $\bar{U} > 0.29$, the hold up problems of the principal and subordinates exactly cancel and subordinates are paid exactly their reservation value $\bar{U}$. If $\bar{U} < 0.29$, however, the principal optimally pays rents in the amount $0.29 - \bar{U}$ to decrease the probability of challenge. Turning to the role of $\bar{V}$, it is easy to calculate that the principal will choose arms-length contracting over the authority whenever $\bar{V} > 0.57$. On other hand, any value of $\bar{U} < S - \bar{V} = 0.43$ would make authority a more efficient organizational form. Thus, for $\bar{U} = 0$, say, the principal optimally chooses an inefficient organizational form.

4.3 Ex ante efficiency: authority vs. arms-length interaction

The literature on organizations often attributes to authority relationships efficiency gains due to their superior ability to resolve coordination problems and make better use of information (Arrow, 1974).

This paper takes these advantages as given, focusing instead on the distortions caused by the cost of exercising authority. In our model, the efficiency gains of authority can be destroyed either through inefficient ex post challenges, or inefficient ex ante choice of arms-length interaction or market exchange. We address the role of renegotiation in alleviating the first distortion later (§4.6); here we focus on the second distortion assuming that no renegotiation is permitted.

Securing subordinates' compliance requires ex ante rents that can distort the principal's choice between authority and alternative arrangements. That is, high rents, which represent the cost of exercising authority, may force the principal to forgo the efficiency gains of an authority relationship. In the model, these gains are $S - \bar{V} - \bar{U} > 0$, while the principal's ex ante payoff from an authority relationship is only $S - \bar{V} - \bar{U} - r$, where $r \geq 0$ represents the ex ante rents needed to secure subordinates compliance. It is easy to generate examples of values of parameters in which $S - \bar{V} - \bar{U} - r < 0$. The contribution of the model is in providing a simple framework linking the primitives of the environment to the likelihood of this distortion.

4.4 Commitment, size, and ex ante efficiency

It is useful to note the efficiency implications of the principal's level of commitment, parametrized by ρ. At one extreme, $\rho = 1$ corresponds to the commitment benchmark (§2.5) whose outcome is efficient. As ρ decreases to 0, the principal's ex ante payoff in an authority relationship, $V(w)$, approaches $-w$ so he strictly prefers not to enter such relationship if either $\bar{V}$ or $\bar{U}$ is strictly positive.[9]

The number of subordinates, N, plays an important role in offsetting the principal's limited commitment and alleviating the resulting distortion. A larger N creates a more severe free-rider problem for subordinates. This follows from the asymptotic results of Rob (1989), Mailath and Postlewaite (1990a) and Ledyard and Palfrey (1994), who show that the probability of

[9]This follows from the fact that $Q \to 0$ as $\rho \to 0$. This is easily seen by noting that as $\rho \to 0$, equation (3) is eventually satisfied for $\bar{h} = 1$, so Q is close to 1.

provision of a public good goes to zero as N increases to infinity.[10] In our setting, as N increases, organizing a collective action by subordinates become more difficult and $Q(w) \to 0$ as $N \to \infty$. Consequently, Program P is (asymptotically) equivalent to the full commitment problem, and all rents to secure compliance eventually disappear.

For fixed (as opposed to asymptotically increasing) N, some degree of collusion is possible. Our analysis reveals how the principal can manipulate collusion through the payment of rents. Paradoxically, subordinates strictly benefit from these manipulations that create impediments to their collective action against the principal. The paradox is due to the fact that the principal's ex ante choice of entering the relationship depends on his expectation of a free-rider problem that impedes subordinates' collusion against him. More formally, without this free-rider problem, the principal optimally sets $e = 0$ and agents are unable to enter into an authority relationship that may pay them strictly positive rents.

Numerical example (2): We verify the effects of increasing the number of subordinates N in the numerical example presented earlier, considering values $N = 100$, 500 and 1000. We find that $\hat{w}$ decreases from 0.29 to 0.20 and 0.15, while the unconditional probability of an effective challenge decreases from 0.14 to 0.03 and 0.02. The ex ante expected value of an authority relationship (assuming $\bar{U} = 0$) improves from 0.57 to 0.77 and 0.83.

Note that this positive effect of N on the principal's payoff uses our particular specification a linear relationship, $\bar{K} = \frac{S}{g}N$, between N and the maximum number of challenges the principal can tolerate. Our model can

[10] In our case, the public good is the launch of an effective challenge, its cost is $\rho\bar{K}w$, and the probability of provision is Q.

be used with alternative specifications of the 'sanctioning technology.' For example, consider the case in which $\bar{K}$ is fixed (independent of N). In this case, the probability of an effective challenge may well increase with N, as illustrated in the numerical example:

Numerical example (3): Maintaining the same parameters as before, but fixing $\bar{K} = 50$, we find that the probabilities of an effective challenge at values $N = 100,\ 500$ and 1000 are $0.14,\ 0.99$ and 0.99 respectively. By contrast with part 2 of the example above, increasing N in this case makes authority less and less attractive.

4.5 Unions and other organized forms of collusion

Our mechanism-design model of collusion can be easily modified to examine the implications of alternative collusive environments. Of particular interest is the role and implications of institutions such as unions, political parties, or social norms.

It is useful to begin with the benchmark of an institution (a union, for concreteness) that faces the same informational and budget constraints as subordinates do on their own. In particular, this institution has no access to subordinates' private valuations nor can it coerce them into participation. Such an institution faces the same free-rider problem as that facing subordinates in the collusive mechanisms of §3, and its role is already covered by our analysis. In this benchmark case, the institution has no independent power or resources; its role is that of a forum through which subordinates can deliberate and coordinate their actions. Interestingly, a union in this case may provide the coordination necessary to implement the optimal collusive mechanism underlying our robustness criterion.

This benchmark sheds light on the role played by subordinates' organization. Our analysis suggests that *any* effect of organized collusive mechanisms (beyond what is already included in our notion of robustness) must take the form of removing or relaxing the constraints we imposed in our description of collusive mechanisms in §3. We briefly examine three such possibilities.

4.5.1 External subsidies

One way in which an organization can facilitate collusion is by breaking the budget balance constraint. The simplest case to consider is that where an organization gives subordinates access to an external source of funding to finance challenges. Specifically, suppose that an amount $\rho \bar{K} w > Z > 0$ is available, then the budget balance constraint becomes:

$$\sum_n c_n(t) \leq Z.$$

Our analysis of optimal mechanisms has a straightforward extension to this case, namely that the optimal mechanism is the same as before, except that it collects a contribution $d(h) = \frac{\rho \bar{K} w - Z}{N}$ from type 1 instead of the $d(h) = \frac{\rho \bar{K} w}{N}$ without subsidies.

A subsidy thus lowers the cost of challenging authority. This makes incentive compatibility easier to satisfy, and consequently lowers the cutoff $\bar{h}$ in equation 3, leading to a higher ex ante probability of challenge and greater rents. The result is that an authority relationship is less attractive to the principal relative to alternative arrangements such as less efficient arms-length contracting or market exchange.

4.5.2 Coercing subordinates into participation

The free-rider problem appears because incentive compatibility and participation require that challengers be compensated (on average) by an amount ρw. An obvious role an organized collusive mechanism, such as a union, can play is coercing subordinates into participating in challenges of authority. Coercion arises in this model through a weakening of the participation constraint.

As a simple example, consider an institution that can apply equal pressure on subordinates to participate in a challenge. This can be formally represented by lowering subordinates' reservation value in the collusive mechanism, so their participation constraint becomes:

$$E[U(t,\sigma,c)|t_n] \geq E[\gamma t_n S|t_n] - z. \qquad (IR - z)$$

Here $z > 0$ measures the extent to which subordinates can be coerced into participation. Under the new constraint (IR-z), the optimal simple mechanisms derived under (IR) can be modified to extract an additional z from each subordinate, collecting a total of $Z = Nz$ which may be viewed as a subsidy as in §4.5.1 above, with identical implications.

4.5.3 Improving transfer technology

Our formulation of collusive mechanisms in §3 assumed that *all* transfer schemes are available to subordinates, implicitly attributing to them a greater ability to commit to such schemes than would be realistic. It is natural to consider a restricted set of mechanisms (as Ledyard and Palfrey (1994) in the public good literature). Such restrictions would impact directly on our analysis in a way that can, at least in principle, be incorporated in our model.[11]

One interesting possibility suggested by our analysis is that the principal may develop impediments to limit subordinates' ability to implement transfers. This divide-and-conquer strategy has a direct benefit to the principal by modifying the collusion constraint 7 so that the probability of an effective challenge $Q(w)$ is lower for every w.

An obvious role for organized collusive mechanisms is that of enforcement and coordination of the mechanism, allowing subordinates access to a fuller set of transfer schemes than would have been possible without such organizations.

4.6 Renegotiation

Challenges in our model are never efficient. Yet a typical solution to Program P will involve some level of challenges, in the sense that the principal will not typically set w so $Q(w) = 0$. Since this outcome is ex post inefficient, it is reasonable to expect some mutually beneficial renegotiation. Here we explore the implications of a variant of the model that include renegotiation.

Specifically, we consider the case in which renegotiation occurs at the beginning of the collusion stage and takes the following form: subordinates agree not to challenge the principal in exchange for a lump-sum payment that they share equally. Clearly, the outcome of such re-contracting will be efficient since the surplus S is generated and none of it is wasted in challenges. Note that anticipation of such re-contracting increases the value of an authority relationship to the principal, and may therefore prevent ex ante inefficiency in the form of the principal choosing inefficient arms-length contracting.

[11] Ledyard and Palfrey (1994) provide such analysis in a model without participation constraints.

To simplify the exposition, it is convenient to assume that the optimal mechanism takes the form $q(h) = 1$ for $h > \bar{h}$ and 0 otherwise.[12] The unconditional probability of challenge in this case is $Q = \sum_{\bar{h}+1}^{N} P(h)$, where $P(h)$ is equal to the probability of h successes in N trials.

From (4), the maximum amount the principal is willing to pay subordinates in exchange of their commitment not to challenge is:

$$A = \sum_{\bar{h}+1}^{N} P(h)[(1-\rho)SN + \rho g\bar{K}], \tag{8}$$

while the minimum subordinates are willing to accept is the ex ante value of the optimal collusive mechanism, which is:

$$B = \sum_{\bar{h}+1}^{N} P(h)[(1-\rho)Sh - \rho\bar{K}w]. \tag{9}$$

The potential surplus achievable by renegotiation is thus:

$$A - B = \sum_{\bar{h}+1}^{N} P(h)[(1-\rho)S(N-h) + \rho\bar{K}(g+w)] \tag{10}$$

How the surplus is divided depends on the bargaining power of the parties, and has no effect on ex post efficiency. On the other hand, the division can affect ex ante efficiency through the principal's decision to enter the relationship. Concretely, consider the extreme case of a principal capable of making take-it-or-leave-it offer, leaving subordinates at their threat point B. The principal now receives total payoff:

$$N(S-w) - \sum_{\bar{h}+1}^{N} P(h)[(1-\rho)Sh - \rho\bar{K}w],$$

a clear improvement over his original payoff of NV.

[12]This underestimates the optimal probability of challenge by at most $P(\bar{h})$, which is small for large N.

Even this lopsided division of surplus does not guarantee ex ante efficiency: relative to the commitment benchmark (in which his payoff is $N(S-w)$), the cost of exercising authority may still be substantial enough to distort his ex ante choice of organizational form. Clearly, less lop-sided divisions of surplus (e.g., a 50/50 division) would lead to even more distortion, and would only reinforce our point.

The result that the ex ante choice of an organizational arrangements may be distorted due to anticipation of subsequent hold-up is in the spirit of the property rights theory of Williamson (1985) and Grossman and Hart (1986). As in these analyses, our explanation is based on an assumption of contractual imperfections (that a_n is not ex ante contractible) and, under renegotiation, the outcome is ex post efficient (no challenge takes place). On the other hand, we do not require the other key ingredients of these theories, namely asset specificity, bounded rationality, and especially the critical role of physical assets. These facts play no role in our analysis, which is based on asymmetric information bargaining and free-riding. We suspect the inclusion of physical assets would not qualitatively alter our analysis.

4.7 Adverse selection of 'trouble-makers' for low w

An interesting consequence of our analysis is the adverse selection effect of low w. Specifically, assume that subordinates learn their types before agreeing to enter the relationship. In this case, their ex ante participation constraint, equation 6, does not bind. The reason is that with $w < \bar{U}$ only type 1 subordinates enter the relationship. In the collusion stage, there is no private information and consequently no free-rider problem. It is then easy to find a collusive mechanism under which subordinates always challenge, and $Q = 1$. The following formalizes this intuition:

Proposition 3 *Suppose that $\bar{V} > 0$. Then if subordinates know their types ex ante then $e = 1$ implies that their ex ante participation constraint, equation 6, does not bind.*

Proof: Consider the Program P with the constraint 6 removed. Only subordinates of type 1 accept a $w < \bar{U}$. Consider the following pivot mechanism: $\sigma_n(t) = a^0$ for $n = 1, \ldots, \bar{K}$, and $\sigma_n(t) = a^*$ for $n > \bar{K}$; $c_n(t) = 0$ for all n and t. It is easy to verify that this mechanism satisfies A.1 and yields $Q = 1$. But then the principal's payoff is non-positive, violating his ex ante participation constraint, equation 5. This implies that if $e = 1$, we must have $w \geq \bar{U}$. ■

Essentially, by setting $w < \bar{U}$ the principal gets an adverse selection problem of hiring only subordinates consisting entirely of 'trouble-makers' eager to challenge him. This adverse selection problem is rather stark in the two type case, but we suspect it will persist (in a milder form) if there are multiple types.

5 Related literature

Our work is motivated by a substantial literature on the central role of authority and reciprocity in firms and organizations. We already mentioned the classic works of Coase (1934), Simon (1951) and Arrow (1974) where authority is largely what *defines* a firm (as opposed to market exchange, say). In Simon (1951) and Arrow (1974), authority consists of an exchange of compliance for wages and restraints in the exercise of authority—although no formal mechanism explaining the terms of this exchange is provided.

The questions addressed in this paper are also related to those in the literature on property rights pioneered by Grossman and Hart (1986). That literature is predicated on the assumption that ownership of physical assets conveys authority and power over subordinates. A common observation is that firms, supported by powerful authority relationships, thrive even when no significant physical assets exist. And even when there are significant physical assets, it is not clear how (and why) their ownership conveys power. In this paper we describe a simple setting where authority arises because of a failure of collective action, independently of any role physical assets may play.

The idea that authority relationships are characterized by reciprocity appears in the literature on gift exchange (Akerlof (1982) and Bewley (1990)) and efficiency wages (see Weiss (1990) for survey). In our model, the give-and-take characterizing authority relationships displays many key features of gift exchanges and/or efficiency wages: subordinates strictly prefer to comply with the principal's authority, even though no formal mechanism exists to enforce it, in exchange for rents above and over their outside options. Here we offer a new rationale for these practices distinct from those found in the literature, with rich comparative static implications. Bewley (1998) provides an extensive survey of employment contracts, supporting of the idea that authority is based on reciprocal exchange in which both parties share mutual obligations.

The ideas presented in this paper are closer to those found in Mailath and Postlewaite's (1991b) and Stole and Zwiebel (1997). Mailath and Postlewaite examine the question of why service firms obtain positive value though their assets consist solely of labor receiving competitive wages. They argue that such firms have value because workers face a collective action problem that prevents them from leaving to form a new firm. There are many key differences with our model; in particular, in Mailath and Postlewaite (1991b) there is no public good problem as we have here (non-participating agents can be excluded from gains), and the principal does not face a trade-off between rents and the probability of collective action.

Stole and Zwiebel (1996) examine how surplus is shared as an outcome of an intra-firm bargaining between a principal and N workers. Roughly, the principal's ability to fire individual workers implies that his bargaining position improves as the number of workers increase. Although there are similarities in the motivation of our work, there are several key differences. Our analysis hinges on the difficulties of collusion among subordinates caused by asymmetric information, while Stole and Zwiebel's model is driven by complete information bargaining between the the principal and individual subordinates. Thus, two key components of our model, the free-rider problem among subordinates and parties' lock-in, have no counter-part in their model. This, in turn, leads to different implications on, say, the role of collusive organizations (such as unions).

The implication of our model that the principal benefits from increasing N is similar to an implication derived by Stole and Zwiebel. It is important to note that, in our model, this conclusion follows from the special specification of a constant critical fraction of challengers, $\frac{\bar{K}}{N}$. As shown by example in Section 4.4, under alternative, equally reasonable assumptions about how $\bar{K}$ varies with N, our model leads to the opposite conclusion, namely that the probability of a challenge to the principal's authority goes to 1. In our framework, these two opposite conclusion are based on the same unifying logic of collusion and free-riding, applied under alternative assumptions about subordinates' environment.

6 Concluding remarks

The motivation for this paper was the observation that, although authority offers enormous advantages as an organizational form, its exercise must entail costs to explain why alternative arrangements such as arms-length contracting and market exchange exist at all. Our answer to this question is that an effective exercise of authority requires that subordinates within an organization be better treated than outsiders. Under this view, the exercise of authority generates natural boundary on the scope of organizations as well as costs for alternative organizational forms—much in the spirit of the classical views of the firm. The paper provides a detailed model that derives the cost of exercising authority from the primitives of the environment.

Our analysis imposed many simplifying assumptions on the environment in order to obtain a tractable, closed form solution. It may be useful to separate assumptions needed for tractability from those driving our analysis. The key idea underlying this paper is the public-good nature of challenging authority as an impediment to subordinates' collusion. Thus, the assumption that all subordinates benefit when the principal capitulates, that subordinates have private information about their types, that there is an irreducible cost to challenging (introduced via $\rho > 0$), and that the principal has limited contractual commitment ($\rho < 1$) are essential for the analysis. On the other hand, the basic results of our model are consistent with generalizations such as multiple actions and types. Finally, the assumption that the principal faces a constant returns to scale technology (i.e. he receives S from every subordinate under his authority) can be weakened considerably.

A Appendix

A.1 Summary of notation

Ex ante stage: S, per capita surplus (§2, p. 4); $\bar{U}$, subordinate's ex ante reservation value (§2.2, p. 5); $\bar{V}$, principal's per capita reservation value (§2.2, p. 5); $e \in \{0,1\}$, decision to enter an authority relationship (§2.5, p. 8); w, upfront reward paid by principal to subordinates (§2.1, p. 4); r, ex ante rents (§4.2, p. 17); $V(w)$, principal's ex ante expected payoff (§4.1, p. 15).

Collusion stage: N, number of subordinates (§2, p. 4); $a_n \in \{a^*, a^0\}$, subordinate n's action (§2.1, p. 4). ρ, measure of principal's commitment (§2.1, p. 5); w', maximum sanction imposed (§2.1, p. 4); $\gamma \in \{0,1\}$, decision to defend authority (§2.4, p. 7); K, number of challengers (§2.4, p. 7); $\bar{K}$, minimum effective number of challengers (§2.4, p. 7). g, net cost of sanctioning a challenge (§2.3, p. 6);

Collusive mechanisms: $t_n \in \{0,1\}$, subordinate n's type (§3.1, p. 9); p, probability of type 1 (§3.1, p. 10); c_n, subordinate n's transfer (§3.1, p. 10); h, number of type 1 subordinates (§3.4, p. 12); $q(h)$, probability of an effective challenge (§3.4, p. 12); Q, unconditional probability of challenge (§4.1, p. 14); (σ, c), a collusive mechanism (§3.2, p. 10);

Simple mechanisms: δ_h, binomial probability (§3.4, p. 13); P, binomial probability (§4.1, p. 14); $\bar{h}$, critical threshold (§2, p. 13); $d(h)$, per capita cost of challenge (§3.4, p. 13);

A.2 Proof of Proposition 1

Many steps in the proof are well-known in the public good literature (e.g. versions of Lemma A.1 appear Mailath and Postlewaite (1990b), and is implicit in Ledyard and Palfrey (1994)). Lemma A.2 borrows ideas from Ledyard and Palfrey. As far as we can see, their results do not directly apply to our setting.[13] This appendix provides a self-contained argument.

Let $i = 0, 1$ represent the subordinate's type, and $h = h(t)$, where $h = 0, \ldots, N$, is the number of type 1 reports in the type profile t. A mechanism is *symmetric* if expected transfers and the probability distribution over actions depend only a subordinate's type i and the total number of reported type 1 subordinates h, but not on the name of the subordinate. Formally, for every type profile t, and corresponding $h = h(t)$, (1) every subordinate of type i receives an expected transfer $c_i(h)$; and (2) a subset of $K_i(h)$ type i subordinates is picked at random (with equal probability) to challenge, where $K_i(h)$ may be random.

Lemma A.1 *Any mechanism (σ, c) satisfying A.1 can be weakly improved on by a symmetric mechanism (σ', c') satisfying A.1 and such that, for every h: (1) $E(U_n|\{t : t_n = 0 \text{ and } h(t) = h\}) = 0$; (2) $K'(t) \in \{0, \bar{K}\}$; and (3) $q(h) = q'(h)$.*

Proof: Consider an arbitrary mechanism (σ, c). Let $\phi : N \to N$ be a permutation of the players' names. Define (σ^ϕ, c^ϕ) so that $\sigma^\phi(t) = \sigma(\phi(t))$ and $c^\phi = c(\phi(t))$, i.e. (σ^ϕ, c^ϕ) applies the original mechanism to the permutation of names ϕ. Let q be the uniform probability distribution on the set of $N!$ permutations of the players' names, and define the anonymous mechanism

[13] The main reasons is that they have no counter part to subordinate actions and they are primarily concerned with settings where IR does not hold.

(σ^q, c^q) to be that in which a permutation ϕ is first chosen at random, then (σ^ϕ, c^ϕ) is applied. Clearly, the probability that an agent challenges and his expected transfer depend only on his type and on the number of type 1 reports h. The probability of provision remains unchanged as

$$E(\gamma|\sigma^a, c^a) = \sum_\phi q(\phi)E(\gamma|\sigma^\phi, c^\phi) = \sum_\phi q(\phi)E(\gamma|\sigma, c) = E(\gamma|\sigma, c).$$

We also note that since the relabeling is not determined by the reports, IR, IC, and BB will continue to hold for all n in any permutation ϕ, and so will hold for uniform probability distributions over such permutations. ■

This allows us to restrict attention to symmetric mechanisms. Next we introduce additional notation. Let $\gamma(h) = q(h)(1-\rho)$ be the probability that the principal blinks given h. Let $\sigma_i(h)$ denote the probability that a subordinate of type i is called on to challenge and $c_i(h)$ the expected transfer when there are h subordinates of type 1.

Under a symmetric mechanism interim expected utility of a subordinate of type i depends only on i, and reduces to:

$$U_i = \sum_{h=i}^{N-1+i} \delta_{h-i}[i\gamma(h)S \underbrace{-(1-\gamma(h))\sigma_i(h)w + c_i(h)}_{y_i(h)}].$$

We next show that IC for type 0 must hold in any symmetric mechanism satisfying BB and IR. Budget balance implies the weaker (per capita) ex ante budget balance: $p\sum_{h=1}^{N}\delta_{h-1}c_1(h) + (1-p)\sum_{h=0}^{N-1}\delta_h c_0(h) \leq 0$. Type 0 IR implies $U_0 = \sum_{h=0}^{N-1}\delta_h(-(1-\gamma(h))\sigma_0(h)w + c_0(h)) \geq 0$. Since $(1-\gamma(h))\sigma_0(h)w \geq 0$ for every h, $\sum_{h=0}^{N-1}\delta_h c_0(h) \geq 0$ and consequently $\sum_{h=1}^{N}\delta_{h-1}c_1(h) \leq 0$. If a type 0 subordinate reports that his type is 1, then his interim utility is

$$U_0' = \sum_{h=1}^{N}\delta_{h-1}(-(1-\gamma(h))\sigma_1(h)w + c_1(h))$$

Since $(1-\gamma(h))\sigma_1(h)w \geq 0$ for every h and $\sum_{h=1}^{N} \delta_{h-1}c_1(h) \leq 0$, $U_0' \leq 0$ while $U_0 \geq 0$ so IC indeed holds.

To satisfy part (1) of the conclusion of the lemma, we modify the mechanism by setting $\hat{y}_0(h) = 0$ for every h and distributing the surplus equally over type 1 (so $y_1(h)$ adjusts by $\frac{(N-h)}{h}y_0(h)$). Note that $y_1(h)$ increases on average, but not necessarily for each h. This ensures that part (1) of the conclusion of the lemma holds. We need to verify that the new mechanism satisfies A.1. By construction, BB continues to hold, and IR for type 0 now holds strictly. Type 1 subordinates' IR and IC constraints continue to be satisfied. IC for type 0 holds because the new mechanism satisfies BB by construction.

Finally, to satisfy part (2), we modify the mechanism by optimizing each σ^{ϕ}, and distributing the proceeds among type 1 agents. For every ϕ modify σ^{ϕ} to $\hat{\sigma}^{\phi}$ as follows: for every type profile t, if $\sigma^{\phi}(t)$ contains less than $\bar{K}$ challenges, then set $\hat{\sigma}_n^{\phi}(t) = a^*$; if $\sigma^{\phi}(t)$ contains strictly more than $\bar{K}$ challenges, set $\hat{\sigma}_n^{\phi}(t) = a^*$ for an arbitrary subset of subordinates, so there are $\bar{K}$ challenges. Modify the transfers, now denoted y_i' for the new mechanism, so that $y_0'(h) = \hat{y}_0(h) = 0$ as before, and $y_1'(h)$ is determined by the equation the budget balance condition $hy_1'(h) + (N-h)y_0'(h) = 0$. Since this (weakly) increases the payoff of type 1 for any h, IR and IC continue to hold for such subordinates. By construction, BB and IR for type 0 continue to hold, so by the previous lemma IC for type 0 also holds. We conclude that the new mechanism continues to satisfy A.1.

Finally, we note that above we only modified the transfers, keeping the probability of launching an effective challenge unchanged, so requirement (3) in the conclusion of the lemma holds. ■

The lemma shows that given the probabilities of the challenges $q(h)$, efficiency pins down the magnitudes of the transfers and the number of challenges. Next we characterize the form of the function $q(h)$ consistent with efficiency.

Lemma A.2 *Every efficient, symmetric mechanism satisfying A.1 is simple.*

Proof: By the last lemma, we may assume that an efficient simple mechanism has transfers of the form $d(h) = \frac{\bar{K}\rho w}{h}$. To simplify notation, write

$$U_1 = \sum_{h=1}^{N} \delta_{h-1} q(h) \left((1-\rho)S - \frac{\bar{K}\rho w}{h} \right)$$

for the interim utility of type 1 when he truthfully report his type, and

$$U_1' = \sum_{h=0}^{N-1} \delta_h q(h)(1-\rho)S.$$

when he mis-report his type to 0. Clearly $\frac{\partial U_1}{\partial q(h)} = \delta_{h-1}\left((1-\rho)S - \frac{\bar{K}\rho w}{h}\right)$ and $\frac{\partial U_1'}{\partial q(h)} = \delta_h(1-\rho)S$, so

$$\frac{\frac{\partial U_1}{\partial q(h)}}{\frac{\partial U_1'}{\partial q(h)}} = \frac{h}{N-h}\frac{1-p}{p}\left(\frac{(1-\rho)S - \frac{\bar{K}\rho w}{h}}{(1-\rho)S}\right) \qquad (*)$$

Consider an h at which $(1-\rho)S - d(h) < 0$. In this case, the ratio of the partial derivatives in (*) is negative, so we may modify the mechanism by reducing $q(h)$ to 0 without violating IC. When $(1-\rho)S - d(h) < 0$ launching a challenge at h is inefficient, so reducing $q(h)$ to 0 also increases subordinates' expected utility. We conclude that, for an efficient mechanism, $q(h) = 0$ unless $(1-\rho)S - d(h) \geq 0$.

To complete the proof of the lemma, suppose that we have an efficient mechanism with the property that there is no $\bar{h}$ satisfying the definition of simplicity. This implies that we can find $h > h'$ such that $q(h) < 1$ and $q(h') > 0$. One may verify directly that (*) is increasing in h whenever $(1-\rho)S - d(h) \geq 0$.[14]

Then we many find $\epsilon, \epsilon' > 0$ such that $\hat{q}(h) = q(h) + \epsilon < 1$ and $\hat{q}(h') = q(h') - \epsilon' > 0$ such that U_1' remains unchanged but U_1 increases. The modified mechanism continues to be incentive compatible, and increases the expected utility of type 1 subordinates, while maintaining that of type 0's constant. We conclude that any efficient mechanism must be simple. ■

Lemma A.3 *Every efficient mechanism satisfying A.1 is simple.*

Proof: Given an efficient mechanism (σ, c), we may use Lemma A.1 to derive a symmetric mechanism that weakly improves on and that preserves the function $q(h)$. The new mechanism must be efficient. Clearly, by the same lemma $K(t)$ is either 0 or $\bar{K}$. Also, by lemma A.2, $q(h)$ has the required property in the definition of a simple mechanism. ■

Proof of Proposition 1: The last lemma proves the first claim of the proposition. It only remains to show that $\bar{h}$ is set as asserted. We first restrict attention to a subset of mechanisms in which we force $q(\bar{h}) = 0$. Let $\hat{h}$ be the smallest h such that $(1-\rho)S - d(h) \geq 0$. By the proof of the previous lemma, $\bar{h} \geq \hat{h}$, and efficiency requires to make $\bar{h}$ as small as possible subject to the incentive constraint for type 1 subordinates.

[14]One may directly calculate the partial derivative with respect to h to be:

$$\frac{(1-p)}{p}\frac{N}{(N-h)^2}\left[1 - \frac{((\bar{K}\rho w)}{(1-\rho)Sh)}\right] + \bar{K}\rho w/((1-\rho)Sh^2)(h/(N-h))]$$

The first expression will be positive if the expression in square brackets is positive, which holds iff $(1-\rho)S >= \frac{\bar{K}\rho w}{h}$.

The incentive constraint for type 1 players requires that

$$\sum_{h=\bar{h}}^{N-1} \delta_h[(1-\rho)S - d(h+1)] \geq \sum_{h=\bar{h}+1}^{N-1} \delta_h[(1-\rho)S]$$

which reduces to (1). By definition of $\bar{h}$, this last inequality is violated for any $h' < \bar{h}$. Thus, any simple with cutoff $h' < \bar{h}$ is not incentive compatible for type 1 subordinates (hence violates A.1). On the other hand, by the arguments above, a simple mechanism with cutoff $h' > \bar{h}$ is not efficient. This establishes the claim that $\bar{h}$ must be set as asserted. ■

REFERENCES

Akerlof, G. A. (1982): "Labor Contracts as a Partial Gift Exchange," *Quarterly Journal of Economics*, **XLVII**, 543-69.

Arrow, K. J. (1974): *The Limits of Organization.* New York: Norton.

Bewley, T. F. (1998): "Listening to Business: A Study of Wage Rigidity," Yale University.

Grossman, S. and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, **94**, 691-719.

Halaby, C. (1986): "Worker Attachment and Workplace Authority," *American Sociological Review*, **51**, 634-49.

Kreps, D. (1990): "Corporate Culture and Economic Theory," in *Positive Perspectives on Political Economy*, ed. by Alt, J. and K. Shepsle. Cambridge: Cambridge University Press, 90-143.

Mailath, G.J. and A. Postlewaite (1990): "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, **57**, 351-67.

Mailath, G.J. and A. Postlewaite (1990): "Workers versus Firms: Bargaining over a Firm's Value," *Review of Economic Studies*, **57**, 369-80.

Ledyard, J. and T. Palfrey (1994): "Voting and Lottery Drafts as Efficient Public Good Mechanisms," *Review of Economic Studies*, **61**, 327-55.

Rob, R. (1989): "Pollution Claim Settlements under Private Information," *Journal of Economic Theory*, **47**, 307-333.

Simon, H. (1951): "A Formal Theory of Employment Relationship," *Econometrica*, **19**, 293-305.

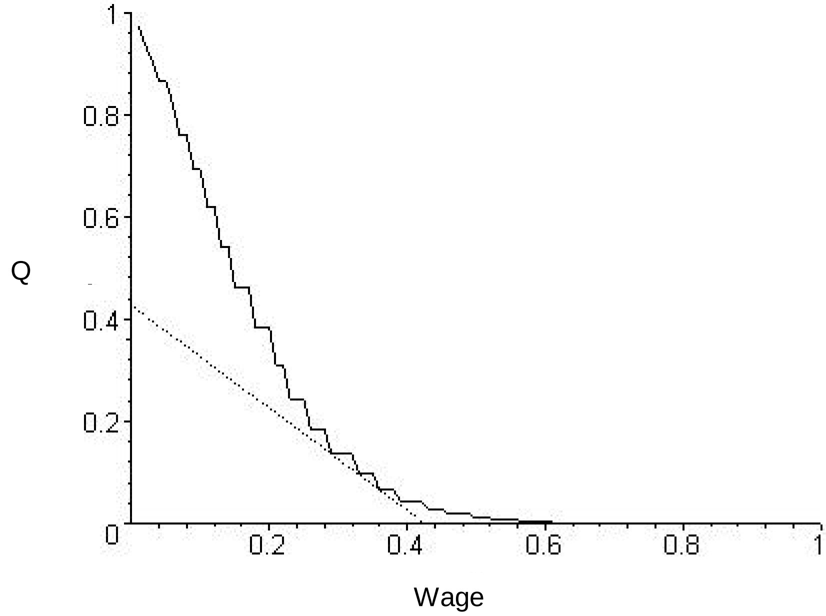
Stole, L. and J. Zwiebel (1996): "Organizational Design and Technology

Choice under Intrafirm Bargaining," *American Economic Review*, **86**, 195-222.

Weiss, A. (1990): *Efficiency Wages: Models of Unemployment, Layoffs, and Wage Dispersion.* Princeton, New Jersey: Princeton University Press.

Williamson, O. (1985): *The Economic Institutions of Capitalism.* New York: Free Press.

Figure 1
The Wage – Challenge Tradeoff*



Challenge Constraint

Isoprofit

*(Parameter values N = 100, S = 1, ρ = 0.5, p = 0.5, g = 2)