ELSEVIER

Journal of Economic Behavior & Organization
Vol. 46 (2001) 165–191

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

A reputational model of authority☆

Nabil I. Al-Najjar*

Department of Managerial Economics and Decision Sciences, J.L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA

Received 16 February 1999; received in revised form 26 July 2000; accepted 07 August 2000

Abstract

The paper provides a model where authority relaionships are founded on reputation. The viability of authority is the result of subordinates' free-riding on each other's challenges, reducing the frequency of challenges, and making reputation worth defending. The party with authority secures subordinates' compliance through the payment of rents to influence the extent of their failure to act collectively and exacerbate the free-rider problem they face. The model provides a framework to explain how the magnitude and form of these rents depend on the primitives of the environment and on the authority's design of its reputation. Applications to employment relationships, dictatorships, and the notion of legitimacy are considered. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: J41; L14

Keywords: Authority relationships; Reputation; Subordinates; Power

1. Introduction

Many group interactions are structured as authority relationships where one party complies with the directives and wishes of another. Employers, bosses, dictators, organizational and political leaders tell others what to do and, more often than not, can expect their directives to be followed without active or overt resistance. While the mechanisms, costs, and benefits involved in such authority relationships remain a puzzle, their pervasiveness is not in doubt. In his seminal work on the employment relationship, Simon (1951) noted that authority is the fundamental characteristic of employment contracts. [1] In fact, the prevailing

☆ Center for Mathematical Studies in Economics and Management Science, Discussion Paper no. 1223.

* Tel.: +1-847-4912-5426; fax: +1-847-467-1220.
E-mail address: al-najjar@mwu.edu (N.I. Al-Najjar).

[1] See also Arrow (1974). This view is also found in sociological studies of employment relationships. Halaby (1986), who elaborates on the central role played by workplace authority, quotes Max Weber: "the hiring of any kind of service for wage or salary ... involves the subjection of the worker under a form of domination" (p. 636).

PII: S0167-2681(01)00192-5

view among economists is that authority over subordinates is what distinguishes interactions within firms from market exchange, [2] and is, therefore, key to understanding organizations.

Authority is also a central theme in other social sciences, where it is viewed as a major force in aligning divergent individual interests into coherent, organized structures (like firms, unions, states and armies). The rise or decline of these organizations is often attributed to the success or failure of their internal authority structures. [3] Perhaps nowhere is authority more evident than in dictatorships, a political form prominent throughout history, where a small group of individuals mobilizes a much larger population into action (or inaction) to serve its interests. In all these examples, who possesses authority is a major determinant of such things as the way interactions are structured, resources are used and wealth is distributed.

It is tempting to explain authority in terms of possession of instruments of power. Variants of such explanations include that a firm's authority over employees stems from its ownership of physical assets, that capitalists control workers because they own the means of production, or that dictators command obedience because they control the guns, and so on. As Arrow (1974) noted, these simplistic explanations are often misleading. Control over instruments of power "cannot be the sole or even the major basis for acceptance of authority" because

> Control mechanisms are, after all, costly. If the obedience to authority were solely due to potential control, the control apparatus would be so expensive in terms of resources used as to offset the advantages of authority (p. 72).

Even when delegated from a more senior authority, or supported by legal rights, such as binding contracts or ownership rights, a viable authority must, at least partly, rest on subordinates' consent. An employer's authority, for instance, would not be viable if he had to spend most of his time and resources implementing disciplinary measures against employees. Indeed, the hallmark of successful leadership is structuring its authority so it is rarely, if ever, challenged. By contrast, the collapse of authority is often linked to a loss of will to face the mounting cost of dealing with an increased frequency of challenges.

That authority relationships ultimately rest on subordinates' consent, on their belief in its power rather than the overt and frequent use of such power, are important aspects often considered the defining features of such relationships. Subordinates' consent does not mean that they like compliance per se, but that they view it as optimal given their expectations about the authority's behavior. Noting the insufficiency of control over instruments of power as basis of authority, Arrow (1974, p. 72) concludes that "authority is viable to the extent that it is the focus of convergent expectations. An individual obeys authority because he expects that others will obey it."

Convergence of expectations as basis for authority poses a puzzle: individuals with conflicting interests would each like to establish authority over others, but only one can ultimately prevail. What makes one party more effective in building expectations about his resolve to defend his authority? What makes expectations converge in this party's favor?

[2] This is a central theme of the transaction cost tradition, following Williamson (1985). Discretion and authority are central to the influence cost view of Milgrom (1988) and Milgrom and Roberts (1982), the property rights approach of Grossman and Hart (1986), and the firm as a carrier of reputation, as in Kreps (1990).

[3] See, for instance, Arrow (1974, p. 65), or Coleman (1990, Chapter 6).

This paper provides a model to address these issues. We consider a stylized environment where a single central player, called the *authority*, interacts repeatedly with N identical long-lived subordinates. The authority's power to direct subordinates is founded on its reputation, which we interpret as expectations about how challenges and compliance will be dealt with. Since subordinates with long-term interest in the relationship may be tempted to challenge authority, the outcome of such contest is decided not just by the players' resources, but also by their expectations about their opponents' behavior. A reputational model of authority should, therefore, account for why one party has greater ability to carry reputation.

In this paper, the authority's main advantage is that it interacts with every subordinate, while subordinates interact only with the authority. This makes it possible to control subordinates by exploiting the *public-good nature* of challenging authority: each subordinate compares the private cost of a challenge with the private benefit of his incremental impact on making the authority blink. Subordinates' free-riding on each others' challenges reduces their frequency and potentially makes authority worth defending.

Free-riding, however, is not sufficient to explain why the authority's reputation dominates. The reason is the considerable arbitrariness of expectation formation in dynamic strategic settings. Maintaining authority seems to require just the opposite of this arbitrariness, namely subordinates' unquestioning submission and their inability to see viable alternatives to compliance. To achieve such a degree of conviction, the authority's reputation should be robust to the uncertain, often arbitrary, nature of the way expectations are formed. We model this by restricting the authority to reputations that are worth defending against any subordinates' behavior that is rational relative to some conjecture they have about the future.

When a reputation is robust in this sense, a compelling logic determines players' expectations, leading to the game's unique equilibrium: the public-good nature of reputation implies that a large fraction of subordinates free-ride, so the authority defends its reputation regardless of what the remaining subordinates do. But if all subordinates are convinced the authority will defend its reputation, they all comply and the actual frequency of challenges drops to zero.[4] Subordinates conclude that compliance is 'the obvious thing to do,' without the need for precise, prior knowledge of others players' actions and expectations.

Thus, the model uses the indeterminacy and arbitrariness of expectations as a source of restrictions on allowable reputations, hence on the authority structures they can support. To put this in perspective, it is useful to contrast the analysis with contracting in static agency settings. Such contracts require that subordinates be left just indifferent between taking the desired action and some other inferior action.[5] The implausibility of this prediction, especially in long-term relationships, is hard to overlook: subordinates left just indifferent between challenging and complying have nothing to lose in the short-run, and might have something to gain in the long-run if their challenges help overturn the authority. But then

[4] This logic departs from the reputation literature pioneered by Milgrom and Roberts (1982), and Kreps and Wilson (1982) in several important ways (see Section 3.6).

[5] See, e.g. Grossman and Hart (1983). Technically, the question is whether the incentive constraint binds. A more detailed discussion appears in Section 3.5.

the authority is exposed to the risk of being challenged frequently, escalating the cost of defending its reputation, and undermining its viability.

By contrast, an authority built on a robust reputation reflects the idea that subordinates view compliance as 'the obvious thing to do.' Such effective focusing of expectations comes at a price, however: to ensure it, the authority must hedge against subordinates' incentives to challenge, even though these challenges never actually occur. In particular, since no free-riding has to occur if subordinates are indifferent between challenging and complying, robust reputations always require that subordinates receive rents for compliance. The authority's problem is then to design its reputation to balance the need to exacerbate subordinates' collective action problem, with the desire to minimize the cost of implementing the necessary rewards and sanctions. The result is a simple framework that links the primitives of the environment to the authority's reputation choice, and to the magnitude and form of the rents needed to secure compliance.

These rents provide a more realistic description of authority relationships, and an insight into the factors responsible for the immense variety of rewards and sanctions used to support them. For instance, extreme sanctions applied to "make an example of someone" are widely practiced in settings where subordinates have no viable outside options. These include some of the more perverse examples of authority relationships, such as dictatorial regimes, Mafia families and insurgency movements.[6] Sanctions involving seemingly gratuitous violence and cruelty are widely documented and, given their cost, would be difficult to explain in the absence of reputational concerns. It is unlikely that subordinates who live under such conditions are left just indifferent between compliance and challenge. Rather, extreme measures are effective means of exacerbating the free-rider problem among potential challengers. This makes it possible for a small group of individuals to control much larger populations using surprisingly small expenditures of resources, and despite the fact that the dominant group could not withstand a collective act of subordinates.

When subordinates have viable outside options, as in employment contexts, the same logic implies that authority is maintained through rewards that make subordinates strictly prefer compliance to challenging. To an outside observer, this might appear to be a form of gift exchange or efficiency wage: employees strictly prefer to work above the contractually specified minimum level, and the employer rewards employees with rents for complying. The efficiency effects of gift exchange and higher wages are well-known.[7] What the model offers is a new and distinct rationale for their use as means of raising subordinates' cost of challenging the employer's authority.

This last point deserves some elaboration: subordinates' collective interest in undermining the authority's reputation does not preclude welfare enhancing authority relationships. In the context of employment, compliance should be interpreted broadly to include such things as refraining from disruptive activities, bargaining, haggling, renegotiation or generally encroaching upon an employer's authority. These are examples of influence activities[8] that

[6] See Gambetta (1993, e.g. pp. 43–46 and 245) for the role of reputation in the operation of Mafia families, viewed as economic enterprises in the business of protection and intimidation. On insurgents' methods of extracting compliance and the role of reputational concerns, see Leites and Wolf (1970, pp. 99–112).

[7] See the surveys by Akerlof (1984), Weiss (1990) and Bewley (1998) for the literature on gift exchange and efficiency wages. More references may be found in Section 4.1

[8] See Milgrom (1988), and Milgrom and Roberts (1982).

are socially wasteful and whose proliferation would be inconsistent with efficiently operating organizations. A credible authority can create efficiency gains that improve individuals' ex ante welfare (before joining the organization) by countering their ex post incentive, as subordinates, to challenge it by engaging in such wasteful activity. Stated differently, one solution to the problem of minimizing influence cost may be to create authority structures that succeed by exploiting subordinates' inability to act collectively. Individuals' ex ante welfare improves if some of the efficiency gains achieved in this manner are distributed through higher wages, better career opportunities and working conditions.

This argument may clarify our assumption of a centralized, star-shaped structure of interaction. This structure, taken here as exogenous, provides a tractable, stylized benchmark that captures key features of authority in important examples like employment relationships, franchising, and centralized political systems. It is natural to ask what forces give rise to such special structure. A commonly made argument [9] is that centralization economizes on information and decision making costs. The argument of the last paragraph provides a separate mechanism to account for the value of centralization through its role in strengthening authority structures.

The model also clarifies why size might matter in establishing and maintaining authority. Size matters not because maintaining authority against a larger number of subordinates is intrinsically more valuable, but because it tends to exacerbate the collective action problem they face. The model, therefore, predicts that discretionary powers to direct resources, resolve disputes and meet new contingencies, will tend to be vested in the larger party. This is broadly consistent with the stylized fact that such powers tend to be vested in employers rather than employees, franchisors rather than franchisees, and so on. [10] Size differences can also account for the asymmetry between the challengers and the authority's incentives to invest in reputation: the authority fully internalizes the benefits from challenging subordinates, so free-riding works in one direction only. The model therefore suggests an explanation for why, in an environment in which incompatible authorities might conceivably arise, the authority of the larger player will tend to prevail.

The free-riding argument implies that the relevant measure of size is the number of *independent* decision-making subordinates. Mechanisms facilitating collusion, such as organizations and other collective bodies, offset the free-riding problem, thus limiting the scope of authority and/or increasing the rents needed to secure compliance. The model, therefore, offers a simple rationale for the time-honored precept of *"divide-and-conquer,"* or the widely documented practice of atomizing individuals and dissolving independent institutions in totalitarian regimes. Dividing opponents need not have any intrinsic value to the authority. Rather, its value stems from increasing the effective number of independent subordinates, and the consequent worsening of their free-riding problem.

The outline of the paper is as follows. Section 2 introduces a static agency benchmark. Section 3 develops a model of a long-term relationship, introduces the notion of robust reputations, and elaborates on its implications. Section 4 discusses examples and applications

[9] For instance, Arrow (1974, p. 68).

[10] Hadfield (1990) reports on the considerable authority of franchisors in franchising relationships. Franchisees may be interpreted as subordinates in the model of this paper. Of course, franchisees retain considerable discretionary power; see the discussion of delegation in Section 5.

to efficiency wages, dictatorships, and legitimacy. Section 5 reviews some of the related literature, and Section 6 concludes.

2. The static agency benchmark

A large player, the *authority*, interacts with N identical subordinates. Subordinate n chooses an action $a_n \in A = \{a^0, a^*\}$, were a^* denotes *compliance*, and a^0 denotes a *challenge*. We may interpret compliance as covering those aspects of performance that can be guaranteed through direct means, such as force, coercion, control over means of production, or compliance with a more senior authority. With this interpretation, the difference between a^* and a^0 reflects aspects of performance that cannot be guaranteed through such direct means.

We assume that subordinates' actions cannot be observed perfectly. This may reflect, for instance, the possibility that subordinates' actions are misinterpreted (by the authority and by other subordinates), or that information about subordinates is gathered through noisy channels, making it impossible to perfectly identify their actions. Formally, assume that a subordinate's action generates a signal $b \in \{b_1, \ldots, b_k\}$ with probability $\pi_a(b) > 0$. Signals are independent given subordinates' actions. The authority chooses a contingent scheme $x = (x_1, \ldots, x_k)$ under which an 'ex post' action $x_k \in \mathbb{R}$ is taken, representing a reward or a sanction in response to signal b_k. The set of all such schemes is denoted X. We make the assumption, reflecting rigidities in the authority's dealing with subordinates, that x must be the same across subordinates (although the ex post actions taken $x(b_n)$ will vary).

Subordinate n's payoff depends on his action and the contingent reward/sanction in the form: $u(a_n) + g(x(b_n))$. The authority's payoff with subordinate N takes the form: $v(b_n) - c(x(b_n))$. The authority evaluates vectors of subordinate actions $(a_1, \ldots, a_N)$ according to the average expected payoff: $1/N \sum_n E[v(b_n) - c(x(b_n)) \| a_n]$.

There is a basic conflict of interest over the subordinate's choice of action: Each subordinate generates a surplus $S > 0$ which accrues (in expectation) to the authority if he complies but that he retains if he challenges. Thus, compliance generates expected benefits $Ev(b\|a^*) = S$ and $u(a^*) = 0$, while challenging generates $Ev(b\|a^0) = 0$ and $u(a^0) = S$.

This single-period interaction follows closely standard agency models (Grossman and Hart, 1983). One key departure is our interpretation of the ex post actions x_k. In standard agency models, these are transfers from the principal to the subordinate. Here we assume, as in agency models, that g is strictly increasing so subordinates prefer higher rewards and dislike harsher punishments. On the other hand, we assume that c is convex with unique minimum at 0, with $c(0) = 0$, indicating that the authority dislikes both rewarding and punishing subordinates. In agency models the ex post actions x_k are transfers, so the principal's preference to make as small a transfer as possible implies a preference to imposing sanctions. Our specification of c captures the idea that sanctions such as reprimands, dismissals, or the elimination of political opponents are costly to implement and represent a net loss to all parties. These examples illustrate our assumptions about c, which implies that the authority would rather obtain compliance without implementing any rewards or sanctions. We shall refer to $x_k > 0$ as a *reward* and $x_k < 0$ as a *sanction*: rewards are transfers to the subordinate, while punishments are costly to both parties.

If the authority were able to offer a binding contract, we have the following benchmark *static agency solution*:

$$\max_{\substack{x\in X\\ a\in\{a^0,a^*\}}} \frac{1}{N}\sum_n E[v(b_n) - c(x(b_n))|a],$$

subject to the incentive constraint:

$$Eg(x\|a^*) - S - Eg(x\|a^0) \geq 0 \qquad \text{(IC)}$$

To avoid trivial cases, we assume that the solution to this problem requires that a^* be implemented. Subordinates may have a viable outside option represented by a (common) reservation value u_0 (as in the case of employment relationships, say) in which case we may also impose the participation constraint:

$$Eg(x\|a^*) \geq u_0, \qquad \text{(PC)}$$

Examples such as dictatorships may be interpreted as settings in which u_0 is so low that it is irrelevant for the solution of this problem. We finally note that the setup already includes a participation constraint for the authority, reflected in the fact that, since it can always pick the trivial scheme x^0 defined by $x^0(b) = 0$ for every b, which guarantees a payoff of zero. Thus, a non-trivial scheme $x \neq x^0$ is chosen only if compliance generates a net benefit i.e. $S - Ec(x|a^*) \geq 0$).

3. Long-term relationships

The static agency setting assumes an exogenously given formal mechanism (e.g. courts) to enforce compliance. We reserve the term 'authorityrel' to situations in which a party commands compliance even though no formal mechanisms exist to enforce it. In this case, compliance is the result of a "convergence of expectations," which we introduce by extending the static model to a long-term interaction between the authority and the same N subordinates. We assume that the authority can commit to meetingshort-term challenges, but cannot commit not to capitulate in the future when facing a large number of challenges. This section models this formally and explores some of the problems arising in an equilibrium analysis of authority.

3.1. Basic setup

Interactions occur at times $t = 1, 2, \ldots$. We also add a stage $t = 0$ that represents a 'reputation design' stage.

3.1.1. Stage $t = 0$

The authority picks a scheme $x^* \in X$, which we interpret as the reputation it follows for the remainder of the game.

3.1.2. Stages $t = 1, 2 \ldots$

A reputation choice x^* in stage $t = 0$ determines an infinite-horizon game $\Gamma(x^*)$ played in periods $t = 1, 2, \ldots$. We define the play of the game recursively as follows: given a history h, the authority chooses $\sigma \in \{0, 1\}$ with $\sigma = 1$ indicating that it defends its reputation (for the current round) by continuing to implement x^*, while $\sigma = 0$ indicates that the authority capitulates to the subordinates. Capitulating is an irreversible act that ends the game; if it occurs, each subordinate collects the total present discounted value of future surpluses $1/1 - \delta S$, and the authority receives 0.[11] If the authority decides to defend its reputation, each subordinate chooses to either comply or challenge — i.e. chooses $a_n \in \{a^0, a^*\}$. This generates a vector of signals $(b_1, \ldots, b_N)$, and a continuation history of the form $h' = (b_1, \ldots, b_N; h)$. In summary, given $\sigma = 1$, the timing of actions and the payoffs are the same as in the static model of Section 2. Players discount streams of payoffs using a common discount factor δ.

Subordinates take their actions knowing the history h and the authority's choice $\sigma \in \{0, 1\}$. Since $\sigma = 0$ means that the game ends, to minimize notation we will refer to subordinates' decisions at h instead of $(h, \sigma = 1)$.

3.1.3. Strategies

A strategy $\vec{\sigma}$ for the authority in $\Gamma(x^*)$ consists of a probability $\sigma(h) \in [0, 1]$ of defending the reputation x^* at each history h.[12] As for subordinates, it is natural to allow them to correlate their actions to reflect the possibility of colluding in their challenges of the authority. For example, subordinates may coordinate by forming unions, political parties, and so on. Our maintained assumption is that such institutions can help subordinates coordinate their actions, but have otherwise no coercive power over them. Formally, we assume that subordinates can, in each stage of the game, use a *correlated strategy*, which we identify with a probability distribution α on the set of pure action profiles $\{a^0, a^*\}^N$.[13] Here, the vector of subordinates' actions $(a_1, \ldots, a_N)$, is drawn according to the joint distribution α. Subordinate N's action a_n maximizes his expected payoff knowing only that the vector of actions of the remaining players, a_{-n}, is drawn at random according to α (but without knowing the actual realization a_{-n}). The standard interpretation of a correlated equilibrium is that subordinates receive correlated private signals on which they may condition their actions. Non-coercive coordination mechanisms, such as unions or social organizations, that provide subordinates with a forum for discussion and coordination but without the power

[11] The analysis is unaffected under less extreme assumptions where subordinates do not appropriate everything. Thus, $\Gamma(x^*)$ is not, strictly speaking, a repeated game, but a dynamic game with a state variable being whether x^0 was ever chosen before. Note that this rigidity appears in the finite-horizon reputation models of Kreps and Wilson (1982), and Milgrom and Roberts (1990) where once the incumbent fails to fight entry, he is revealed to be weak and the outcome of the game unravels by backward induction. There, the posterior on the incumbent's type acts as a state variable; here we build this rigidity directly in the specification of the game.

[12] Thus, a subordinate is sure that x^* will be implemented in the current period, but may still believe that his actions today could undermine the authority's resolve to defend its reputation in the future. This gives the authority a slight edge because it can make a short-term commitment to x^*.

[13] The formal definitions of correlated strategies and equilibria are the standard ones — see Appendix A. See Myerson (1992) for more elaborate treatment and motivation. Note that Nash equilibrium is the special case in which the joint distribution α is uncorrelated.

to coerce them, may be interpreted as the correlation devices that give rise to correlated equilibria.

Finally, we restrict attention (as commonly done in the literature) to symmetric strategies: every action has the same ex ante probability of being played by all subordinates (the formal definition is in Appendix A). Let $\mathcal{A}$ denote the set of *symmetric correlated strategies*. A subordinates' strategy $\vec{\alpha}$ in $\Gamma(x^*)$ assigns to every history h a symmetric correlated strategy $\vec{\alpha}(h) \in \mathcal{A}$ played at h.

3.1.4. Continuation values

Fix a history h, and suppose subordinates expect the continuation play $(\vec{\sigma}, \vec{\alpha})$. Let $U(h)$ denote the present discounted expected payoff of a subordinate when $(\vec{\sigma}, \vec{\alpha})$ is played. Our assumption of symmetric strategies guarantees that if subordinates have the same expectations about the continuation play $(\vec{\sigma}, \vec{\alpha})$, as they would in any equilibrium, then their continuation values $U(h')$ must also be the same.

3.2. Collective versus private incentives

Consider a history h at which subordinates hold expectations that a profile $(\vec{\sigma}_h, \vec{\alpha}_h)$, with value $U(h')$, will be played in every continuation h'. If subordinate N believes a profile a_{-n} will be played, then compliance by this subordinates is incentive compatible if

$$Eg(x^*|a^*) - S - Eg(x^*|a^0) \geq \delta[E(U|a_n = a^0) - E(U|a_n = a^*)]. \qquad (\mathrm{IC}')$$

The LHS represents the short-term trade-off already captured by the static incentive constraint (A.1). What is new is that RHS reflects this subordinate's belief about his *influence* on the authority's future behavior. This concept of influence, which appears in a wide range of strategic settings, is formally defined and developed in Al-Najjar and Forman (1999). Note that (IC′) reduces to the static constraint (A.1) if either subordinate N is myopic ($\delta = 0$), or the continuation values do not depend on which action he takes.

The dynamic incentive constraint illustrates the *public-good nature* of challenging authority: the LHS of (IC′) represents the private benefit from compliance, while the RHS reflects potential future benefits from challenging authority. Since other subordinates also benefit from changes in the continuation value, (IC′) underestimates the benefits of challenging by not taking into account the collective interest of the remaining $N - 1$ subordinates. Challenging authority here is a public good in which subordinates under-invest. Our analysis focuses on how the authority manipulates this public good problem to its advantage, and the constraints that such manipulations impose.

3.3. Robust reputations

Subordinates' incentives to comply, reflected in the dynamic incentive compatibility constraints (IC′) above, clearly hinge on their expectations about the continuation values $U(h')$. We introduce a concept that formalizes the idea that subordinates' expectations are rationalized as part of an optimal response to some 'theory' about the authority's future behavior.

We shall assume that, at any history h, subordinates (as they do in any equilibrium): (1) form expectations about the behavior of the authority $\vec{\sigma}_h$ at each continuation history h';

and (2) behave optimally given these expectations, in the sense of choosing a best response to $\{\vec{\sigma}_{h'}\}$. Formally, $(\alpha, \{\vec{\alpha}_{h'}\})$ is a best response to $\{\vec{\sigma}_{h'}\}$ if $\vec{\alpha}_{h'}$ is a best response to $\vec{\sigma}_h$, for every h', and α is a (correlated) equilibrium of the stage game with payoffs augmented by continuation values $\delta U(h')$.[14] We call α *locally optimal* at h if there is $\{\vec{\sigma}_{h'}\}$ and $\{\vec{\alpha}_{h'}\}$ such that $(\alpha, \{\vec{\alpha}_{h'}\})$ is a best response to $\{\vec{\sigma}_h\}$. In this case we say that $\{(\vec{\sigma}_{h'}, \vec{\alpha}_{h'})\}$ (or the corresponding continuation values $U(h')$) *rationalize* α. Let $\mathcal{A}(x^*) \subset \mathcal{A}$ denote the set of locally optimal profiles.

Definition. x^* is robust if it is a strictly dominant strategy against every $\alpha \in \mathcal{A}(x^*)$; we let $X^* \subset X$ denote the set of robust reputations.

A robust reputation hedges against any subordinates' behavior that can be rationalized as optimal against *some* theory about the authority's behavior. We require subordinates to form a common expectation about the future behavior of the authority and behave optimally given these expectations, as they always do in any equilibrium.[15] Equilibrium requires, in addition, that expectations are correct. We discuss this in greater detail later.

It is useful to note that the trivial scheme x^0 is always robust.[16] Thus, the existence of a robust reputation is not an issue in this model. The more interesting question is whether there exist *non-trivial* robust reputations $x^* \in X^* - \{x^0\}$. We explore this issue in Sections 3.5–3.7.

Let $\vec{\sigma}^*$ denote the strategy of never capitulating (i.e. $\vec{\sigma}^*(h) = 1$ for every h) and let $\vec{\alpha}^*$ be the subordinates' strategy in which all subordinates comply at every h. As we discuss in Section 3.8, robustness as a criterion on a reputation x^* is intermediate between requiring $\vec{\sigma}^*$ to be an equilibrium of $\Gamma(x^*)$ — which only requires optimality relative to the actual behavior of subordinates — and requiring $\vec{\sigma}^*$ to be a strictly dominant strategy — which requires optimality against any behavior of subordinates. In fact, robustness implies that compliance is the unique equilibrium of $\Gamma(x^*)$:

Proposition 1. *For any non-trivial robust reputation* x^*, *the profile* $(\vec{\sigma}^*, \vec{\alpha}^*)$ is the unique equilibrium of $\Gamma(x^*)$.

The intuition for the proof is as follows (all proofs are in Appendix A). If $(\vec{\sigma}, \vec{\alpha})$ is an equilibrium, then subordinates' behavior at a history h must be locally optimal (i.e. $\vec{\alpha}(h) \in \mathcal{A}(x^*)$) since their behavior must be rationalized by the equilibrium continuation values. Note that this does not rule out that some subordinates challenge in that round. In Appendix A we show that if x^* is robust, then $\vec{\sigma}^*$ is strictly dominant strategy against

[14] In a best response, every subordinate optimizes given the strategies followed by the authority and other subordinates. Note that best responses are defined relative to *all* unilateral deviations, not just those consistent with symmetric correlated profiles.

[15] In particular, subordinates' expectations may differ from the actual strategy followed by the authority, and may be inconsistent across periods.

[16] To see this, in the game $\Gamma(x^0)$ subordinates suffer no sanctions and earn no rewards for complying, so any expectation they might hold about the future play of the authority would require them to best respond by playing $a_n = a^0$ in every period. Consequently, $\mathcal{A}(x^0)$ consists of singleton profile in which every subordinate challenges. Clearly x^0 is strictly dominant strategy against $\mathcal{A}(x^0)$, and x^0 is, therefore, robust.

any strategy $\vec{\alpha}$ such that in $\vec{\alpha}(h) \in \mathcal{A}(x^*)$, so $\vec{\sigma} = \vec{\sigma}^*$. Given this, subordinates conclude that challenging is pointless, so $\vec{\alpha}^*$ is their unique, in fact strict, response to the authority's behavior σ^*, so we must also have $\vec{\alpha} = \alpha^*$.

The key point is that robustness implies not just the uniqueness of equilibrium, but also a mechanism explaining how expectations are formed, and why subordinates find compliance 'the obvious thing to do.' In light of the proposition, one may interpret robustness as a requirement ensuring that compliance is a *focal point*, in the sense of ensuring that subordinates conclude that it is the optimal course of action, based only on the fundamentals of the game and without precise knowledge of the authority's strategy as would be required in equilibrium.

3.4. Reputations design

Observed authority relationships display an immense variety in terms of the intensity of the rewards and sanctions used. This variety presumably reflects differences in the players' objectives and the constraints imposed by the environment in which they interact. Here, I provide a simple framework accounting for an authority's choice of reputation as a function of the primitives of its environment (i.e. g, c, N, δ and so on).

The idea is that the authority's designs its reputation x^* in anticipation of the outcome of its subsequent interaction with subordinates in $\Gamma(x^*)$. Specifically, consider the constrained maximization problem:

$$\max_{\substack{x\in X\\ a\in\{a^0,a^*\}}} \frac{1}{N}\sum_n E[v(b_n) - c(x(b_n))\|a]$$

subject to the robustness constraint: [17]

$$x \in \mathrm{cl}(X^*). \tag{R}$$

The robustness condition (R) says that the authority hedges against a wide range of outcomes in the game $\Gamma(x^*)$.

Proposition 2 below shows that any non-trivial robust reputation must satisfy the static incentive constraint, which is why (A.1) is redundant in the problem above. Furthermore, as shown in Lemma A.2 in Appendix A, any such x^* must satisfy $S - Ec(x^*|a^*) \geq 0$, so the authority's share of the surplus, which is S minus the expected cost of the rewards and sanctions it implement, must be non-negative (this may be viewed as a participation constraint for the authority).

3.5. Rents and the viability of the static agency benchmark

A well-known result in static agency models is that the optimal contract should leave subordinates just indifferent between compliance and challenge, and between participating

[17] $\mathrm{cl}(X^*)$ denotes the closure of the set of robust reputations $X^* \subset \mathbb{R}^K$. Thus, we are allowing the use of reputations that can be approximated arbitrarily close by robust ones. This is necessary because we use *strict* dominance in the definition of robustness.

in the relationship or choosing their outside option (Grossman and Hart, 1983). A contract for which either the incentive or participation constraint does not bind can be modified to increase the principal's expected payoff without violating these constraints. The following proposition says that robust reputations involve positive rents. In particular, the optimal static agency contract cannot be supported by a robust reputation against subordinates with a long-term interest in the relationship.

Define the *rents for compliance* $G(x)$ under x as the difference between what a subordinate gets by complying rather than challenging: $G(x) = Eg(x|a^*) - S - Eg(x|a^0)$.

Proposition 2. *For any $\delta > 0$ and any non-trivial robust reputation $x \in X^* - \{x^0\}$, $G(x) > 0$. In particular, (A.1) does not bind and subordinates strictly prefer compliance to challenge.*

These rents are paid even though, when x^* is robust, $\Gamma(x^*)$ has a unique equilibrium in which none of the dynamic incentive constraints actually bind. More formally, at any history h, the continuation value in the unique equilibrium is $(\delta/1-\delta)Eg(x^*|a^*)$, independently of the action taken by any subordinate at h. This implies that each subordinate's influence (i.e. the LHS of (IC′)) vanishes and the dynamic incentive constraint reduces to the static agency benchmark (A.1). Proposition 2 then says that (A.1) is not binding under any non-trivial robust reputation.

Why cannot the authority squeeze more surplus by lowering the rents it pays to subordinates — just like the principal does in a static agency relationship? The answer lies in the fact that robustness requires the authority to hedge against any subordinates' behavior that can be rationalized with respect to *some* theory about how the game unfolds, not just the theory that happens to be part of a putative equilibrium postulated by the modeler. Robustness, therefore, reflects the authority's uncertainty about the determinants of subordinates' expectations. In the model, this translates into the requirement that the authority takes into account dynamic incentive constraints not just relative to the equilibrium continuation values, but also relative to values generated by best responses to subordinates' non-equilibrium expectations about how the game unfolds.

In summary, the reasoning underlying the static agency model breaks down because in a long-term relationship, leaving subordinates just indifferent between complying and challenging means they suffer no cost from challenging, yet might gain if their challenges lead to a better continuation as a result of undermining the principal's reputation.

3.6. Effects of size and the discount factor

Next we turn to the effect of the number of subordinates N and their discount factor δ on the magnitude of rents paid. Turning first to N, write $X^*(N)$ to denote the set of robust reputation as a function of the number of subordinates N.

Proposition 3. *For any $\delta > 0$ and $\epsilon > 0$ there is N and $x^* \in \{X^*(N) - x^0\}$* such that $G(x^*) < \epsilon$.

Increasing N worsens the free-rider problem among subordinates because the fraction of subordinates who believe they are pivotal enough to justify the cost of challenging decreases.

Robustness becomes less and less binding, and the optimal reputation choice approaches the static agency contract.

The effect of the discount factor δ is similar: as subordinates become more myopic (i.e. as $\delta \to 0$), the RHS of (IC′) vanishes so this constraint reduces to the static (A.1) constraint. The point is that rents are paid only to control the incentives of subordinates with long-term interest in the relationship, so the motive for these rents disappears as they become more myopic. In summary, we have a continuity result in which increasing N and/or decreasing δ exacerbates the free-riding problem, leading to a reputation choice approaching the static agency contract.

It is useful to contrast our results with the reputation models in Milgrom and Roberts (1982), and Kreps and Wilson (1982). There, an incumbent faces a sequence of myopic challengers. Since these challengers have no long-term incentive to test the authority's resolve, their behavior is completely driven by short term incentives which are fully captured by the static incentive constraint (A.1); in particular, challengers always free-ride and the public-good nature of reputation is trivial. The role of reputation is then simply as a device that replaces binding contracts (unavailable because, say, an incumbent cannot write a contract which commits him to fighting entry). The role of reputation in Milgrom and Roberts (1982), and Kreps and Wilson (1982) is that of replicating what the static agency contract (had it been available) would have achieved. In many settings of interest, including contracting settings such as employment or franchising, models with short-run opponents may not be appropriate. The treatment of this paper reflects the problems arising when subordinates' long-term stake in the relationship is as substantial as that of the authority, and hence, their incentives cannot be reduced to short-term considerations.

3.7. Pivotalness and the characterization of robust reputations

To derive a characterization of the set of robust reputations, we restrict attention to correlated strategies in which a mixed profile $(\alpha_1, \ldots, \alpha_N)$ is drawn and announced publicly to all players.[18] Fix x^*, and suppose that the continuation values $U(h')$ rationalize the profile α at stage h, and that subordinate N challenges with positive probability (that is, α_n assigns positive weight to a^0). Incentive compatibility implies that for this subordinate to challenge, his influence on the continuation values must exceed the cost of challenging:

$$[E(U|a_n = a^0) - E(U|a_n = a^*)] \geq \underbrace{\frac{1}{\delta}[Eg(x^*|a^*) - S - Eg(x^*|a^0)]}_{\frac{1}{\delta}G(x^*)}. \qquad (*)$$

The LHS represents subordinate n's influence on the continuation values (i.e. the extent to which his choice of action changes expected continuation value), while the RHS is the net present value of all future rents paid for compliance. Inequality ((∗)) says that to induce a subordinate to challenge, his influence on the outcome must exceed the threshold $(1/\delta)G(x^*)$, which may be thought of as his private cost of challenging. Call subordinate

[18] Note that in this case the set of correlated equilibria reduces to the convex hull of the set of Nash equilibria.

N pivotal (relative to α, x^* and U) if $(*)$ holds. Clearly, the more subordinates can be made pivotal, the greater the number of challenges the authority might face.

Define

$$K(x^*, \delta) = \max_{\alpha \in \mathcal{A}(x^*)} \#\{n \text{ such that } (*) \text{ holds}\},$$

to be the maximum number of subordinates that can be convinced they will be pivotal over all profiles $\alpha \in \mathcal{A}(x^*)$ that can be rationalized by some expectations about how the the game unfolds. Al-Najjar and Smorodinsky (2000) provide a general characterizations of influence and the maximal number of pivotal players which we use to prove the following.

Proposition 4. *$K(x^*, \delta) = N$ if $G(x^*) \leq 0$. Otherwise there is $\bar{K}$ such that $K(x^*, \delta) = N$ when $N < \bar{K}$, and $K(x^*, \delta) = \bar{K}$ when $N \geq \bar{K}$.*

Intuition: when $(1/\delta)[Eg(x^*|a^*) - S - Eg(x^*|a^0)] = 0$, subordinates receive no rents, hence have nothing to lose by challenging. It is then easy to find expectations that induce every subordinate to challenge. Rents, on the other hand, raise subordinates' opportunity cost of challenging, requiring that their influence on the continuation values be large enough to offset these higher costs. The proposition says that the number of subordinates that can be made pivotal is bounded by a $\bar{K}$ that does not depend on N. If $N < \bar{K}$, this bound has no force, but as N increases most players are non-pivotal under any locally optimal profile.

Proposition 4 implies that the maximum fraction of challengers $K(x^*, \delta)/N$ decreases rapidly to 0 as N increases. On the other hand, if a fraction $\gamma \in [0, 1]$ of subordinates challenge, defending x^* nets $(1 - \gamma)[S - Ec(x^*|a^*)] + \gamma Ec(x^*|a^0)$. Thus,

$$\gamma(x^*) = \frac{S - Ec(x^*|a^*)}{S - Ec(x^*|a^*) - Ec(x^*|a^0)}$$

is the maximum fraction of challenges the authority can tolerate indefinitely and still defend its reputation. Lemma A.2 shows that $0 < \gamma(x^*) < 1$ whenever $x^* \in X^* - \{x^0\}$.

The above observations can be used to characterize the robustness of a reputation in terms of the relationship between the maximal fraction of challenges subordinates could launch in an incentive compatible fashion, $K(x^*, \delta)/N$, and the maximum fraction of challenges the authority can tolerate indefinitely:

Proposition 5. *A reputation $x^* \neq x^0$ is robust if*

$$\frac{K(x^*, \delta)}{N} < \gamma(x^*)$$

and only if

$$\frac{K(x^*, \delta)}{N} \leq \gamma(x^*).$$

The proposition provides a mapping that determines (up to closure) the set of robust reputations X^* in terms of the primitives of the environments. Thus, given x^*, the critical

threshold $\gamma(x^*)$ is completely determined by the S, g, and c, while the maximal fraction of challengers K is determined by N, δ, c, g. The proposition then can be used to test whether x^* is robust.

Finally, the last two results clarify the role of our assumption that subordinates imperfectly observe each other's actions (expressed as the requirement that $\epsilon > 0$). Without this assumption, subordinates play a repeated game with observed actions, whose set of subgame equilibria includes an equilibrium in which: every subordinate challenges on the equilibrium path with probability 1; if one subordinate ever fails to challenge, then all subordinates comply in all future periods. This equilibrium corresponds to the implausible social behavior in which the decision of *each* individual subordinate to comply leads every other subordinate to switch from challenging to complying. Roughly, in this equilibrium *every* subordinate is pivotal in determining the outcome of the game, and no free-riding need to occur. [19] This implausible outcome, first identified by Green (1980), and Sabourian (1990) in the context of repeated games, is eliminated by introducing imperfect observability of the subordinates' actions.

3.8. Expectations and reputation: discussion

Our analysis hinges on authority being based on *robust incentives*, formalized as the robustness requirement that reputation choice satisfies $x \in \mathrm{cl}(X^*)$. In formulating this restriction we required that: (1) subordinates hold a common expectation about the authority's future behavior, and (2) they play a best response to these expectations. Note that *any* equilibrium, by definition, must satisfy (1) and (2). On the other hand, an equilibrium further requires that subordinates' expectations coincide with the authority's true future behavior.

In the absence of formal mechanisms to enforce compliance, authority is largely founded on expectations. Assuming at the outset that expectations coincide with the truth typically leads to expectational bubbles that can be used to support a wide range of outcomes as equilibria of the model. For instance, subordinates may comply because they expect the authority to defend its reputation; the authority finds it optimal to do so because subordinates comply. The problem is that the opposite conclusion can be supported by a different expectational bubble: subordinates challenge because they expect the authority to blink, which is optimal because subordinates challenge. In both cases expectations are based on self-referential, self-confirming reasoning, rather than a mechanism which generates these expectations from the fundamentals of the subordinates' environment. If we take seriously the idea that authority is founded on expectations, these examples suggest that equilibrium analysis begs the very question of the source of authority.

Robustness rules out such circular reasoning by providing a mechanism for generating expectations that is independent of the equilibrium ultimately played. In the model, subordinates' beliefs about the future consequences of their actions are derived from common (but otherwise arbitrary) expectations about the authority's future behavior and the assumption that they behave optimally given these expectations. Since expectations are

[19] Formally, in Proposition 4, if $\epsilon = 0$, then $K(x^*, \delta) = N$ regardless of the magnitude of rents, $G(x^*)$, which means that one can always find outcome functions that make every player pivotal.

not assumed to coincide with the true play, the circularity described above does not arise.

Our criterion of robustness requires an ex ante choice of reputation x^* such that the strategy $\vec{\sigma}^*$ of subsequently defending it is intermediate between the two extremes of being part of an equilibrium and being a strictly dominant strategy. Since equilibrium model of behavior implies knowledge of the opponents' strategies, requiring $\vec{\sigma}^*$ to be part of an equilibrium of $\Gamma(x^*)$ implicitly assumes that the authority believes it can make accurate predictions about subordinates' subsequent behavior. As mentioned earlier, this is too weak a requirement, since an equilibrium model of $\Gamma(x^*)$ does not explain the source of subordinates' expectations. Robustness has a bite because it requires x^* to be optimal against a broader range of subordinates' behaviors and expectations. At the other extreme, a stronger robustness criterion would require a reputation choice x^* such that $\vec{\sigma}^*$ is strictly dominant strategy in $\Gamma(x^*)$. Such criterion would provide a more viable foundation for authority since the play in $\Gamma(x^*)$ would not hinge on subordinates' expectations or whether they best respond to these expectations. Unfortunately, strict dominance is too stringent a requirement in a rich strategic context like the one considered here: no reputation x^* can ever be robust in the stronger sense that $\vec{\sigma}^*$ is optimal against *any* behavior of subordinates. Our notion of robustness weakens this by requiring that x^* is strictly dominant only against the class of locally optimal subordinates' behaviors. Local optimality assumes subordinates best respond to common expectations, but drops the assumption that these expectations necessarily coincide with the truth.

4. Applications and examples

The predictive content of the model derives from the link it establishes between the exogenously given, observable features of the environment, and the authority's optimal choice of reputation. As a test of the plausibility of the model's implications, I provide three examples of how alternative authority structures might arise as a result of optimal reputation choice.

4.1. Gift exchange and efficiency wages

Neoclassical as well as agency theoretic predictions of employment compensation have long been criticized for their failure to reflect important aspects of real-world employment relationships. One set of stylized facts, reported by Bewley (1998), is that successful management treats employees better than the minimum, contractually agreed-upon level, and employees reciprocate by contributing effort, care and dedication in excess of their contractually specified levels. This practice, often reflected by concepts like morale, gift exchange, efficiency wages, and worker attachments [20] is particularly significant in view of the fact that most employment contracts are highly incomplete, so the explicitly specified dimensions of performance are, indeed, quite limited.

[20] See Akerlof (1982) for exposition of the main idea of morale and gift exchange, Weiss (1990) for a survey of the literature on efficiency wages, Halaby (1986) on worker attachment, and Bewley (1998) for a more detailed account of current theory.

The model of this paper provides a new and distinct explanation for this practice. The employer offers salaries and work conditions above the minimum level (an efficiency wage) in exchange for employees' voluntary compliance with his authority. To see how this follows from the model, interpret a^0 and x^0 as the minimum, contractually specified levels of compliance and compensation, respectively. Proposition 2 implies that the use of a robust reputation to support this exchange requires subordinates to strictly prefer compliance to challenging.

The predictive content of the model is in accounting for the magnitude and form of rents as a function of the primitives. Giving employees 'something to lose' in the form of an efficiency wage is, a priori, not the only way to provide incentives; coercion — giving subordinates 'something to fear' — could also make them strictly prefer compliance. How can the model account for the use of inducement rather than coercion? A special feature of employment relationships is the availability of a valuable outside option (workers can simply quit their jobs to seek employment elsewhere). Limited liability rules and other legal restrictions on the severity of sanctions represent additional exogenous restrictions on employers' ability to resort to sanctions to extract compliance. These special features of employment relationships, not shared by other authority structures such as dictatorships, imply that rents for compliance must take the form of additional rewards, resulting in the failure of the participation constraint to bind.[21] Several suggestions appeared in the literature to explain why workers are treated better than the minimum, contractually-specified level. Some are based on the need to overcome monitoring and screening costs (Weiss, 1990), while others revolve around employees' sentiments towards the firm, and the positive psychological effects that a better treatment can have on workers' morale. The present model is not inconsistent with these explanation. Rather, it provides a complementary, and to my knowledge new, mechanism that leads to the payment of efficiency wages. This mechanism shows that qualitatively similar, if not observationally indistinguishable, predictions about efficiency wages and gift exchange may be a consequence of the use of reputation by welfare-maximizing individuals' to structure long-term interactions.

An interesting test of the model's prediction is provided by the traditional distinction between primary and secondary labor markets: "Primary sector jobs have stability, low quit rates, good working conditions, promotion according to a promotion ladder, acquisition of skills, and good pay. In contrast, secondary sector jobs have high quit rates, harsh discipline, little chance of promotion, low acquisition of skills, and poor pay" (Akerlof, 1984, p. 79). Higher turnover in secondary sectors means lower expected duration of stay with any particular employer, hence a lower subordinates' discount factor δ. The difference between these two sectors provides a natural comparative static test of the model's predictions as subordinates' long-term interest in the relationship, measured by δ, vary. Lower δ weakens the robustness constraint ($\bar{R}$), so optimal reputation choice approaches that in a static agency setting, where subordinates have no long-term interest (see Section 3.6). Rents for compliance, which may correspond to higher monetary pay, better working conditions and better treatment on the job, consequently disappear as δ approaches 0.

[21] In the notation of the model, attractive outside options correspond to higher values of u_0, while exogenous restrictions on the use of sanctions may be captured by a g function that rises rapidly with the severity of sanction (negative values of x).

The model's prediction of a negative correlation between the magnitude of rents and turnover rate appear broadly consistent the distinction between the primary and secondary sectors above.

Finally, the practice where parties voluntarily refrain from extracting the maximum private benefit from an exchange are pervasive not just in employment contracts and personnel management, but in nearly every other long-term social and political interaction where conflicting reputational concerns arise. It is folk wisdom that leadership and authority over subordinates with 'nothing to lose' is difficult, if not impossible, to maintain. Greed, in the form of milking subordinates for all they are worth, is not just morally questionable, but is often taken as sign of bad leadership. [22] The model of this paper suggests that rents for compliance are a fairly general and robust consequence of using reputation as a foundation for authority relationships.

4.2. Dictatorships

Dictatorship is one of the most common forms of government throughout history. [23] Yet their existence and, often, remarkable stability raise several puzzles. The only resource under the direct physical control of the dictator is the ability to issue directives — cheap talk, with little or no direct consequence on subordinates' actions and welfare. What gives force to these directives is subordinates' willingness to carry them out, presumably due to a "convergence of expectations" in the dictator's favor.

Dictatorships are not monolithic organizations. It is physically impossible for a single individual to control a large population directly, so dictatorships inevitably rely on layers of subordinates charged with controlling and monitoring lower layers. The stability of the dictator's rule depends on the distribution of rents for compliance throughout the hierarchy.

At the lowest level there are the masses of (typically) unorganized population. Although the resources available to a regime are usually trivial compared to the potential resources of the population, a large N implies that free-riding is rather severe. Individuals behave nearly myopically, allowing control of a large population by a much smaller group. The role of size in simplifying control of large populations has been noted by observers of dictatorial regimes. Commenting on Stalin's rule, Michael Polanyi wrote:

> The stability of such naked power increases with the size of the group under its control, for a disaffected nucleus which might be formed locally by a lucky crystallization of mutual trust among a small number of personal associates would be overawed and paralyzed by the vast surrounding masses of people whom they would assume to be still loyal to the dictator. Hence it is easier to keep control of a vast country than of the crew of a single ship in mid-ocean (quoted in Wrong (1979, p. 94)).

The only countervailing force in a large population is the presence of mechanisms that organize and coordinate subordinates, such as independent institutions (opposition parties,

[22] A well-known quote of Alexander Solzhenitsyn illustrates this point: "You only have power over people so long as you don't take everything away from them. But when you've robbed a man of everything, he's no longer in your power — he's free again."

[23] See Wintrobe (1998), who provides an insightful analysis of the subject.

religious institutions, and so on). Consequently, breaking or weakening organized resistance, "divide-and-conquer" strategies, and atomizing the population are, of course, the hallmark of many dictatorial regimes. Such practices exacerbate free-riding by increasing the number of independent decision-making subordinates.

A more subtle issue is the extensive use of sanctions rather than rewards. Using the model, it is easy to see that rewards are intrinsically more costly: conditional on compliance, a^*, sanctions are rarely applied, while rewards would have to be awarded frequently. Furthermore, unlike employment relationships, subjects of a dictatorship have little alternative but to live under the regime, so their outside options are limited. Clearly, limitations on the dictator's ability to impose sanctions always exist, varying widely from case to case and providing a useful comparative static test of the model's predictions. Taking all these factors into account, the model predicts that sanctions (to the extent feasible) dominate rewards, at least at the population level.

The model also suggests a possible reason for an often noted stylized fact that rewards relative to sanctions increase at higher levels of the hierarchy (e.g. Wintrobe, 1998). The key factor here is that the number of subordinates shrinks at higher levels in the controlling hierarchy. While the number of (potentially) pivotal subordinates K does not necessarily change, their fraction relative to members of that layer, K/N, increases rapidly as N shrinks. Each subordinate can potentially have a large impact, so rents must correspondingly increase. But why increase rewards rather than sanctions? A subtle issue which the model points to is the potentially destabilizing effect of sanctions: they increase the short-term cost of challenging, but also increase the potential gain from undermining the authority's reputation. This should be contrasted with the unambiguous effect of rewards, which raise the short-term as well as long-term cost of challenging. This asymmetry between rewards and sanctions leads to an increase in positive inducements as one moves up the hierarchy.

Finally, the analysis illustrates the model's ability to generate a wide range of authority structures. In Wrong (1979), classification of authority relationships, gift exchange and efficiency wages are examples of *authority by inducement*, where a promise of reward is offered in exchange for compliance. Dictatorships are an example of *coercive authority*, using force "to establish credibility, and thus, to create a future power relation based on the threat of force that precludes the necessity of overt resort to it" (p. 41). Reputation plays critical role: force is "used less for the immediate effects on its victims than to establish or maintain a relation of coercive authority in the future" (p. 42). The model explains that the use of reputation as foundation for authority is capable of generating a variety of authority structures.

4.3. Legitimacy and the boundaries of authority

A central theme of the political science and sociology literature on authority is that submission to authority is never absolute, but limited by rules delineating the scope of activities over which authority may *legitimately* be exercised.[24] Simon's (1951) notion of

[24] In his study of authority in organizations, Scott (1998) defines: "authority is legitimate power." Similar, if less stark, statements can be found throughout the sociology literature.

acceptance set of activities over which an employer can exercise discretion, and Arrow's (1974) emphasis on the need for restrictions to ensure responsible exercise of authority, are very much in this spirit.

Although the boundaries delineating the scope of authority are rarely sharply drawn, they nevertheless represent real constraints on what an authority can and cannot do. For instance, it may be legitimate for an employer to exercise discretion in assigning production tasks to employees, but not in asking them to wash his car or provide sexual favors. By shaping subordinates expectations, legitimacy consequently has considerable practical importance: collapse of authorities is often linked to subordinates' perception to their loss of legitimacy, while their stability is attributed to acquiring or retaining such legitimacy.

Legitimacy raises several complex and multi-faceted issues whose treatment is beyond the scope of this paper. However, a feature common to all notions of legitimacy is the presence of restrictions on the scope of activities over which authority may be exercised. My goal here is limited to show how the model explains the role of reputation in accounting for this key feature of legitimacy.

Imagine that subordinates' behavior consists of choosing levels of L separate activities that may be relevant to the authority. For instance, one such activity may be a worker's performance on a production task while another represents idiosyncratic, personal aspects of his conduct. As in Section 2, we may also interpret the activities as compliance with state-contingent, discretionary directives. Identify an action a_k with a bundle of (levels) of these activities. For simplicity, assume that there is a finite set of such actions $A = \{a_0, a_1, \ldots, a_k\}$, with a^0 continuing to represent challenge, while $a_1, \ldots, a_k$ represent various types of compliance. For instance, different a_k's might represent different subsets of the L basic activities over which subordinates comply.

The reputation design problem now has two components. First, for any given possible reputation a_k, we can find the optimal scheme $\tilde{x}_k$ and corresponding rents G_k needed to support it. This is just the reputation design problem of Section 3.4, with a^* replaced by a_k. A second, new, dimension is the authority's ability to choose among the various levels of compliance.

As a benchmark, assume that a_k is the action that would have been implemented in the optimal static agency contract,[25] under which the subordinate is left with no rents. If a_k is the only level of compliance open to the authority, then the analysis of Section 3 implies that positive rents $G_k > 0$ must be paid to implement it. These rents could be substantial; G_k might be so high that no robust reputation can implement it.

A richer set of actions $\{a_1, \ldots, a_k\}$ ameliorates the authority's predicament. By allowing for the option of a lower level of compliance, $a^* = a_k$, in exchange for a lower rent, G_k, the authority can establish a reputation in situations which it might not otherwise have been able. Note that lower levels of compliance may reflect restrictions on the number of dimensions over which authority may be exercised.

The model, therefore, generates restrictions on the scope of authority based on this trade-off between the desire to achieve a high level of compliance and minimize the rents necessary to implement it. Since this trade-off depends on observables, such as the number

[25] Or equivalently, using a reputation against myopic subordinates.

of subordinates and the strength of their long-term interest in the relationship, it can potentially generate useful, non-vacuous predictions about observed authority structures. For instance, relative to the setting with myopic subordinates, a more limited scope of authority arises in long-term relationships.

5. Related issues and literature

Despite their pervasiveness in economic interactions, and despite Simon's (1951) early emphasis of their role in employment contracts, there has been relatively little formal analysis of authority relationships in competitive models of exchange, or in models of contracting under asymmetric information. One reason is that control over subordinates in these models is provided through exogenously given, formal mechanisms, like the rules governing exchange in a market, or complete contracts enforced through a third party. As noted by Arrow (1974), the need for authority arises precisely when such mechanisms fail to satisfactorily perform allocative tasks in organizations.

The major departure from this tradition is the transaction cost literature (e.g. Williamson, 1985). This literature emphasizes interactions where well-defined property rights and complete contracts which anticipate all possible contingency are unavailable, thus forcing the issue of what enables one party to direct the actions of others. Prominent in this tradition is Grossman and Hart's (1986) model of ownership rights over the use of physical assets as a source of authority.[26] Asset ownership in their model does not convey direct control over subordinates, whose right to control their own actions is inalienable. Rather, ownership gives indirect authority through its effect on the set of options available once new contingencies unfold. Closer to the spirit of this paper is Kreps and Wilson's (1982) model in which the rights of control and the discretionary powers they entail are founded on reputation. While closer in spirit, there are too many differences for a meaningful comparison. I further comment on Kreps' paper in Section 6.

Subordinates in many organizations, such as in franchising and employment settings, retain considerable discretionary power in directing day-to-day operation. Different organizational structures may be thought of as reflecting alternative patterns of internal delegation of authority. The focus in this paper is on the source of authority. The issue of delegation is fundamental to understanding how authority is structured, and is obviously complementary to studies of its source.

The concept of authority is often associated with that of discretion, namely the right to direct subordinates' actions as new contingencies unfold. The employer's authority in Simon (1951), for instance, enables him to pick, ex post, an action at his own discretion. The present model is consistent with the exercise of discretion if we interpret compliance as "obeying the authority's directives", where these directives may be state-contingent. While discretion is not inconsistent with the model, our focus will be on achieving compliance rather than investigating the nature of discretionary power conveyed by authority. It should

[26] This idea is also part of the Marxist paradigm that capital ownership is the source of domination of the working class. See Selznick (1969, pp. 65–67) for the role of property rights in Marxist ideology.

also be noted that, while closely linked, discretion and authority are distinct concepts which can be usefully treated separately. [27]

Two other related models are Chwe (1998), who study the coordination problem facing subjects in rising up to authority. Many of the intuitions presented by Chwe are related those developed in this paper. However, the modeling approaches are complementary, as Chwe's emphasis is on the role of the network structure of subordinates' interaction, while this paper emphasizes the free-riding problem in a repeated setting. Another paper is Al-Najjar and Forman (1999) who study authority in employment relationships in a two stage game with observed actions. Their model lacks the explicit dynamic aspect emphasized in this paper and assumes that actions are observable (hence, does not allow for moral hazard). On the other hand, Al-Najjar and Forman allow subordinates to have private information and explore more elaborate collusive schemes among subordinates, modeled there as direct revelation mechanisms.

By contrast with the economics tradition, questions about the source of authority, who acquires it and how it is maintained have been central topics of research in other social sciences. The political science and sociology literatures examine authority in relation to a host of other concepts, such as power, leadership, discretion, and legitimacy. Uniform, generally accepted definitions of these concepts and the links between them are, unfortunately, lacking. Wrong's (1979) classic essay, *Power*, provides a comprehensive account of the literature on authority and power. Authority relationships are also discussed extensively in Coleman (1990), as well as in the literature on organizations and employment relationships (e.g. Selznick, 1969; Halaby, 1986; Scott, 1998).

Sociologists draw a distinction between authority and power. While both involve the ability to direct subordinates to take actions which may be contrary to their immediate self-interest, one of the defining features of authority relationships is that subordinates' compliance is mostly voluntary or consensual. Wrong (1979, p. 37), who asserts this view, quotes Weber: "every genuine form of domination implies a minimum of voluntary compliance, that is, an interest . . . in obedience." See also Coleman (1990, p. 71). One contribution of the present paper is to provide a mechanism explaining why subordinates form expectations under which voluntary compliance is optimal, and what price has to be paid to secure such compliance.

Finally, sociology and political science studies often invoke reputation and free-riding in informal arguments about authority. For instance, in discussing authority within revolutionary movements, Coleman (1990, p. 482) criticizes the view of such organizations as monolithic entities because, from the perspective of the insurgents, "revolution is a public good, and, like any public good, it gives rise to the free-rider problem." See also Wintrobe (1998, p. 106) discussion of dictatorships. This paper's contribution relative to these informal insights is the idea that subordinates' failure to act collectively is the result of an artificial public good problem, one created by the authority for the purpose of manipulating

[27] Discretion may be achieved by means other than authority, e.g. through an enforceable contract stating that the subordinate will do what his superior tells him to do. Here, the superior has discretion supported not by direct authority over the subordinate, but by whatever source of power used to enforce the contract. An authority relationship may also exist without significant scope for discretion: what a dictator wants from his subjects may be as simple as: do not revolt. In this case, establishing authority need not involve discretion.

their incentives. This makes explicit the authority's incentive to design its reputation to exacerbate the subordinates' free-riding problem. The closer connection between reputation and free-riding makes it possible to address questions about how the primitives shape the scope of authority, and the form and magnitude of the rents needed to support it.

6. Concluding remarks

When reviewing the argument that authority stems from control over the instruments of power, Arrow noted that this is "not a sufficient explanation of obedience to authority even at the immediate level ... [nor is it] ... a sufficiently deep one."[28] This paper provides some insight into the role of reputation as a basis for authority relationships and as an explanation of how these relationships are structured. Free-riding is suggested as a unifying principle that can account for parties' different abilities to carry reputations and to sort out the maze of resources and constraints underlying authority relationships. Analytically, the paper provides a model in which inherently dynamic phenomena, like reputation and authority, can be captured by a robustness constraint added to a standard static agency model.

Several important questions about authority relationships, of course, remain. One important issue is raised by Kreps and Wilson (1982) view of corporate culture, embodied in a firm's reputation, as the 'principle' followed by the firm in such things as exercising discretion over subordinates and adjudicating disputes. The principle is not an explicit, state-contingent rule, but a general guide to be interpreted and re-evaluated as new, previously unforeseen, contingencies unfold. Applied to the present model, the legitimacy of an authority over its subordinates consists of establishing, and subsequently adhering to, a set of principles that delineates the range of activities and subordinates' actions over which its power legitimately extends. The present model captures restrictions on the scope of authority reflected in explicitly stated rules (Section 4.3), but not restrictions embodied by general principles of the sort Kreps has in mind. Modeling such principles is considerably more difficult, apparently for the same sort of reasons that unforeseen contingencies and incomplete contracting are difficult to model. I hope to pursue this in future work.

Acknowledgements

I am indebted to Greg Greiff, Chris Forman, Ramon Casdesus-Masanell, Jim Dana and Daniel Diermeier. I also thank Tim van Zandt, Jeff Ely, Ray Deneckere, Peter Klibanoff, Tim Feddersen, Marc Ventresca, Brian Uzzi, and seminar participants at Northwestern, Wisconsin, Michigan, Rochester, and Windsor. All remaining errors are my own.

Appendix A. Proofs

A.1. Some formal definitions

For completeness, I provide the (standard) formal definitions missing in the description of the model. Aumann's concept of correlated strategy is introduced here in the stage game

[28] Arrow (1974, p. 71).

in which each subordinate N chooses $a_n \in \{a^0, a^*\}$. Formally, a correlation device is a probability space (Ω, Σ, P) and σ-algebras $\Sigma_n \subset \Sigma$ for each N. Here, Ω encompasses all uncertainty, including players' randomizations, and Σ_n is subordinate N's information about the realized state. A strategy for player N is a function $\alpha_n : \Omega \to \{a^0, a^*\}$ measurable with respect to his information Σ_n. The collection of strategies $(\alpha_1, \ldots, \alpha_N)$ induce a joint distribution on the set of pure action profiles $\{a^0, a^*\}^N$. Wecall $(\alpha_1, \ldots, \alpha_N)$ *symmetric* if $P\{\omega : \alpha_n(\omega) = a^0\}$ does not depend on N.

A profile $(\vec{\sigma}, \vec{\alpha})$ induces a distribution on sequences of actions and signal realization which may be used to compute present discounted expected payoffs in the standard way. We define a correlated best response at a history h given continuation values $U(h')$ for all continuation histories $h' = (b_1, \ldots, b_N; h)$ as follows. For notational simplicity, let a_{-n} and b_{-n} denote the vectors of actions and signals of subordinates other than N. The expected payoff of strategy α_n to subordinate n can be expressed as the sum of two components. First we have

$$E \sum_{b_n} \pi_{\alpha_n(\omega)}(b_n) g(x(b_n)) \tag{A.1}$$

where expectations are taken over ω. Second, subordinate N receives a payoff in terms of the continuation values $U(h')$:

$$\delta E \sum_{b_1, \ldots, b_N} \left[\prod_{n=1}^{N} \pi_{a_n}(b_n) \right] P(a_n, a_{-n} | \Sigma_n)(\omega) U(b_1, \ldots, b_N; h | a_n, a_{-n}) \tag{A.2}$$

where $P(a_n, a_{-n} \| \Sigma_n)(\omega)$ represents this subordinates' assessment of the actions taken by others conditional on his information. Expected payoff to the strategy α_n is the sum of (A.1) and (A.2) above; and α_n is best response if there is no other strategy α_n' that yields higher expected payoff.

A.2. Proofs

Notational conventions: We write $\vec{\alpha} \in \mathcal{A}(x)$ if $\vec{\alpha}(h) \in \mathcal{A}(x)$ for every history h. Unless indicated otherwise, h' will denote the immediate successor history to h.

Lemma A.1. $1/1 - \delta[S, Eg(x^*|a^*)]$ *is the smallest interval containing the range of present discounted expected subordinate's payoff of profile* $(\vec{\sigma}h, \vec{\alpha}h)$ *ranging over all* h, N, *and* $\vec{\alpha} \in \mathcal{A}(x^*)$.

Proof. Clearly, by complying in each period, any subordinate can guarantee himself an average continuation value of at least $(1/1 - \delta)Eg(x^*|a^*)$, so this is a lower bound on the continuation values. This bound is achieved by taking $\vec{\sigma}_{h'} = \vec{\sigma}^*$ for all h'. In the other direction, any conjecture $\vec{\sigma}_{h'}$ consists of a (possibly state-contingent) sequence of a^0 and a^* choices, against which a subordinate's stage payoff is an average of S, $Eg(x^*|a^*)$, and $S - Eg(x^*|a^0)$. By our assumptions on stage-game payoffs, such convex combinations are bounded above by S. This bound is achieved by taking the conjecture that the authority chooses x^0, to which subordinates' best response guarantees them S each period. □

Proof of Proposition 4. Consider first the case of x^* such that $G(x^*) < 0$. For any h, taking the continuation play of the authority to be $\vec{\sigma}^*$, subordinates' unique best response is the profile $\vec{\alpha}^0$ in which the profile α^0 where each chooses action a^0 with probability 1 is played at every history. This implies that $U(h')$ is constant so continuation values have no effect on short-term incentives. Then $G(x^*) < 0$ implies that subordinates optimal strategy at h is α^0, so $K = N$ as required. The proof for the case where $G(x^*) = 0$ is similar, except now it is no longer *strictly* optimal for subordinates to challenge. In fact, any correlated profile is in $\mathcal{A}(x^*)$, in particular one can find continuations that rationalize $K = N$.

Assume now that $G(x^*) > 0$. Fix a history h and a strategy $\vec{\alpha} \in \mathcal{A}(x^*)$. Let H' denote the set of next period histories, and define the function $d : H' \to [0, 1]$ by $d(h') = U(h')/S - Eg(x^*|a^*)$, so the function d takes values in the interval [0,1]. Playing $\vec{\alpha}(h)$ implies that a profile of mixed actions $(\alpha_1, \ldots, \alpha_N)$, $\alpha_n \in [0, 1]$, is drawn and made public (note that the profile need not be symmetric, and that randomizations are independent). Write $Ed(b_{-n}, b_n|a_{-n})$ to denote the expected normalized continuation value d fixing subordinate N's signal to be b_n and calculating expectation on b_{-n} given that other subordinates play a_{-n}. Define

$$Z_n(\alpha_{-n}) = \max_{b_n} Ed(b_{-n}, b_n \| \alpha_{-n}) - \min_{b_n} Ed(b_{-n}, b_n | a_{-n}).$$

where Z_n is the maximum difference subordinate N's signals can have on the expected value of d given this subordinate's uncertainty about other subordinates' signals — but knowing their mixed actions. By Theorem 1 in Al-Najjar and Smorodinsky (2000), for any constant $r > 0$, there is an integer, $K(r)$ such that there can be at most $K(r)$ subordinates for whom $Z_n > r$, uniformly over all functions d (hence, all continuation values U), profiles $(\alpha_1, \ldots, \alpha_N)$, and N (for this theorem to hold, we need each signal b_n has positive probability, i.e. $\pi_{a_n}(b_n) > 0$ which is guaranteed by assumption).

Abusing notation, set $K(x^*) = K(G(x^*))$. That is, $K(x^*)$ is the number of subordinates who would have the incentive to challenge under x^* had they been able to control their signal — but remain uncertain about the signal realization of others. Since subordinates control only the distribution of their signals, not the signals themselves, $K(x^*)$ remain also an upper bound on the number of challengers, as required. Taking the least upper bound yields the result. □

Lemma A.2. *For any non-trivial robust reputation $x^* \in X^* - \{x^0\}$, we have $S - Ec(x^*|a^*) > 0$. For any such reputation, $\gamma(x^*) \in (0, 1)$.*

Proof. Since x^0 guarantees a non-negative payoff, for x^* to strictly dominate x^0 it must yield a strictly positive payoff. To prove the second assertion, we note that our assumption that $c > 0$ except at $x_k = 0$, and $x^* \neq x^0$ imply that $Ec(x^*|a^0) > 0$, hence $\gamma(x^*) \in (0, 1)$. □

Proof of Proposition 2. From the proof of Proposition 4, $G(x^*) = 0$ implies $K = N$ so the necessary condition for robustness in Proposition 5 cannot hold. □

Proof of Proposition 3. We show that any x^* for which (IC) strictly holds must belong to X^* for large enough N. Theorem 1 in Al-Najjar and Smorodinsky (2000) implies that for

any $r > 0$, $K(r)$ is bounded independently of N. Thus, for any such x^*, $K(x^*)/N \to 0$ as $N \to \infty$. Since $\gamma(x^*) > 0$ is constant independent of N, Proposition 5 implies that $x^* \in X^*$ for large enough N. Choosing $x^* \in X^*$ arbitrarily close to static agency solution and noting that the latter guarantees 0 rents establishes the claim. □

Proof of Proposition 5. First suppose that $K/N < \gamma(x^*)$. Then in any $\alpha \in \mathcal{A}(x^*)$ no more than K subordinate challenges. From the definition of γ it is strictly dominant strategy to play $\sigma = 1$, hence x^* is robust. To prove necessity, suppose that $K/N > \gamma(x^*)$ then there is $\alpha \in \mathcal{A}(x^*)$ in which K subordinates challenge. Clearly playing x^* is not optimal against such α, and x^* cannot be robust. □

Lemma A.3. *If x^* is a non-trivial robust reputation, then $\vec{\sigma}^*$ is strictly dominant strategy against any $\vec{\alpha} \in \mathcal{A}(x^*)$.*

Proof. As a notational convention, we write $\vec{\sigma}' \succ_{\vec{\alpha}} \vec{\sigma}$ if the authority strictly prefers (in terms of present discounted payoffs) strategy $\vec{\sigma}'$ to $\vec{\sigma}$ when playing against $\vec{\alpha}$. Note that $\succ_{\vec{\alpha}}$ is continuous relative to weak convergence of strategies. [29]

Fix $\vec{\alpha} \in \mathcal{A}(x^*)$. We show that $\vec{\sigma}^* \succ_{\vec{\alpha}} \vec{\sigma}$ for any $\vec{\sigma} \neq \vec{\sigma}^*$. Starting with $\vec{\sigma} \neq \vec{\sigma}^*$, define the new strategy $\vec{\sigma}_1$ as follows. Take a shortest-length history h with the property "$\vec{\sigma}(h) < 1$, but $\vec{\sigma}(j) = 1$ for all histories j preceding h" (that is, h is a 'first' history at which x^* is not played with probability 1 under $\vec{\sigma}$). Define $\vec{\sigma}_1(h) = 1$; for any immediate successor h' to h, if $V(h') < 0$ set $\vec{\sigma}_1(h') = 0$; and let $\vec{\sigma}_1$ coincides with $\vec{\sigma}$ otherwise.

Then $\vec{\sigma}_1 \succ_{\vec{\alpha}} \vec{\sigma}$. To see this, note that the only change in values occur at histories starting with h, since $\vec{\sigma}_1$ and $\vec{\sigma}$ coincide otherwise. At h, the changes in payoff can be decomposed into the current period change, and the change in the expected continuation values. By construction, the continuation values $V(h')$ at successor histories h' either stayed the same, or strictly increased (this occurs for h' with $V(h') < 0$). [30] On the other hand, the assumptions that x^* is robust and $x^* \neq x^0$ imply that playing x^* instead of x^0 at h generates strictly positive short-term gain. Thus, switching to $\vec{\sigma}_1$ strictly increases payoffs at h, weakly increases them at all successor histories h', and leaves them unaffected otherwise. Consequently, $\vec{\sigma}_1 \succ_{\vec{\alpha}} \vec{\sigma}$.

If $\vec{\sigma}_1 = \vec{\sigma}^*$, then we are done. Otherwise, repeat the process above, with $\vec{\sigma}$ replaced by $\vec{\sigma}_1$, proceeding lexicographically relative the length of histories at which $\sigma \neq 1$. This generates a new strategy $\vec{\sigma}_2$ such that $\vec{\sigma}_2 \succ_{\vec{\alpha}} \vec{\sigma}_1$. In doing so, continuing in this manner, if we ever obtain a strategy $\vec{\sigma}_k = \vec{\sigma}^*$, then we are done, otherwise we have an infinite sequence $\{\vec{\sigma}_k\}$ that converges to $\vec{\sigma}^*$ weakly as $k \to \infty$ and has the property $\vec{\sigma}_{k+1} \succ_{\vec{\alpha}} \vec{\sigma}_k$, for all k. By continuity of $\succ_{\vec{\alpha}}$, $\vec{\sigma}^* \succ_{\vec{\alpha}} \vec{\sigma}_k$ for all k, hence $\vec{\sigma}^* \succ_{\vec{\alpha}} \vec{\sigma}$. □

[29] A sequence of pure strategies $\{\vec{\sigma}_m\}$ weakly converges to a strategy $\vec{\sigma}$ if for every integer K, there is M large enough such that, $m \geq M$ implies that $\vec{\sigma}_m$ and $\vec{\sigma}$ coincide on all histories of length at most K. This guarantees that discounted expected payoffs converge for any discount factor $0 \leq \delta < 1$.

[30] This is the only place where the assumption that once the authority blinks it permanently loses its reputation is used. The assumption implies that the continuation value following the play of x^0 is 0, ensuring an unambiguous comparison between x^* and x^0.

Proof of Proposition 1. Let $(\vec{\sigma}, \vec{\alpha})$ be an equilibrium. Clearly $\vec{\alpha} \in \mathcal{A}(x^*)$. The last lemma shows that $\vec{\sigma}^*$ is strictly dominant against $\vec{\alpha}$, hence, we must have $\vec{\sigma} = \vec{\sigma}^*$. Given $\vec{\sigma}^*$, the subordinates's unique best response is $\vec{\alpha}$, so $\vec{\alpha} = \vec{\alpha}^*$. □

References

Akerlof, G.A., 1982. Labor contracts as a partial gift exchange. Quarterly Journal of Economics 47, 543–569 .

Akerlof, G.A., 1984. Gift exchange and efficiency-wage theory: four views. American Economic Review, Papers and Proceedings 74, 79-83.

Al-Najjar, N.I., Forman, C., 1999. Reciprocity and the costs of authority relationships. MEDS Department, Kellogg GSM, CMSEMS Working Paper, Northwestern University.

Al-Najjar, N.I., Smorodinsky, R., 2000. Pivotal players and the characterization of influence. Journal of Economic Theory 92, 318–342.

Arrow, K.J., 1974. The Limits of Organization. Norton, New York.

Bewley, T.F.,1998. Listening to Business: A Study of Wage Rigidity. Yale University, New York.

Chwe, M.S.-Y., 1998. Communication and coordination in social networks. Review of Econimic Studies, in press.

Coleman, J.S., 1990. Foundations of Social Theory. Belknap Press.

Gambetta, D., 1993. The Sicilian Mafia: The Business of Private Protection. Harvard University Press, Cambridge.

Grossman, S., Hart, O., 1983. An analysis of the principal-agent problem. Econometrica 51, 7–45.

Grossman, S., Hart, O., 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. Journal of Political Economy 94, 691–719.

Green, E., 1980. Noncooperative price taking in large dynamic markets. Journal of Economic Theory 22, 155–182.

Hadfield, G., 1990. Problematic relations: franchising and the law of incomplete contracts. Stanford Law Review 42, 927–993.

Halaby, C., 1986. Worker attachment and workplace authority. American Sociological Review 51, 634–649.

Kreps, D., Wilson, R., 1982. Reputation and imperfect information. Journal of Economic Theory 27, 253–279.

Kreps, D., 1990. Corporate culture and economic theory. In: Alt, J., Shepsle, K. (Eds.), Positive Perspectives on Political Economy. Cambridge University Press, Cambridge, pp. 90–143.

Leites, N., Wolf Jr., C., 1970. Rebellion and Authority: An Analytic Essay on Insurgent Conflict. The RAND Corporation, Santa Monica, California.

Milgrom, P., Roberts, J., 1990. Bargaining costs, influence costs, and the organization of economic activity. In: Alt, J., Shepsle, K. (Eds.), Positive Perspectives on Political Economy. Cambridge University Press, Cambridge, pp. 57–89.

Milgrom, P., 1988. Employment contracts, influence activities and efficient organization design. Journal of Political Economy 96, 42–60.

Milgrom, P., Roberts, J., 1982. Predation, reputation and entry deterrence. Journal of Economic Theory 27, 280–312.

Myerson, R., 1992. Game Theory, Analysis of Conflict. Harvard University Press, Cambridge.

Sabourian, H., 1990. Anonymous repeated games with a large number of players and random outcomes. Journal of Economic Theory 51, 92–110.

Scott, W.R., 1998. Organizations, 4th Edition. Prentice-Hall, New Jersey.

Selznick, P., 1969. Law, Society, and Industrial Justice. Russell Sage Foundation, New York.

Simon, H., 1951. A formal theory of employment relationship. Econometrica 19, 293–305.

Weiss, A., 1990. Efficiency Wages: Models of Unemployment, Layoffs, and Wage Dispersion. Princeton University Press, Princeton, NJ.

Williamson, O., 1985. The Economic Institutions of Capitalism. Free Press, New York.

Wintrobe, R., 1998. Political Economy of Dictatorship. Cambridge University Press, Cambridge.

Wrong, D.H., 1979. Power: Its Forms, Bases, and Uses. The University of Chicago Press, Chicago.