

# Labor Associations: The Blue Wall of Silence<sup>☆</sup>

David K. Levine<sup>1</sup>, Andrea Mattozzi<sup>2</sup>, Salvatore Modica<sup>3</sup>

---

## Abstract

We provide a model showing that a firm's employees may prefer to protect shirkers because this optimally reduces overall effort. This is the case when labor demand is inelastic and individual behavior is easily monitored. In this case, employees have a strong incentive to conceal information about peers' performance from firms, what has been infamously known as the *blue wall of silence* in the case of the police.

*Keywords:* Labor Associations, Monitoring Costs, Self Organizing Groups.

---

---

<sup>☆</sup>First Version: March 7, 2021. We gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL\_01.

*Email addresses:* [david@dklevine.com](mailto:david@dklevine.com) (David K. Levine), [andrea.mattozzi@eui.eu](mailto:andrea.mattozzi@eui.eu) (Andrea Mattozzi), [salvatore.modica@unipa.it](mailto:salvatore.modica@unipa.it) (Salvatore Modica)

<sup>1</sup>Department of Economics and RSCAS European University Institute and Department of Economics WUSTL

<sup>2</sup>Department of Economics European University Institute

<sup>3</sup>Università di Palermo, Italy, Department of Economics, Business and Statistics (SEAS)

## 1. Introduction

*Baltimore cop, stripped of police powers after fatally shooting unarmed teen, kept on payroll for 28 years.*

Baltimore Brew August 12, 2021

Why do unions often protect the worst workers, that is those who are so bad that they clearly are not a majority? An obvious context evoked by the news article mentioned in the incipit is that of the police in the USA, but teachers unions are often blamed for similar practices.<sup>4</sup> Our explanation for this apparent puzzle is that it has to do with discouraging effort provision. We provide a model showing that under some economically intuitive conditions, employees indeed prefer to shield shirkers because this reduces overall effort. Furthermore, they have a strong incentive to prevent the flow of information about employees' performance from firms, what has been infamously known as the *blue wall of silence* in the case of the police.

In our model there are two types of employees: workers and shirkers. Workers prefer to provide effort because they get satisfaction out of a job well done. Shirkers prefer not to provide effort. Consider first the case in which only employees have information about their peers' effort. If shirkers outnumber the workers (a formal condition will be given below) then it will be advantageous to all employees, regardless of their type, to allow workers to work and provide incentives for shirkers to work harder. If instead workers dominate then it will be advantageous to all employees, regardless of their type, to allow shirkers to shirk and provide incentives for workers to work less hard. In the former case we show that in equilibrium labor demand must be elastic and in the latter case inelastic.

Our theory says that labor associations will protect their weakest members when demand for labor is inelastic both without enforcement (which we do not observe) and in equilibrium (which we do observe). Given the ubiquity

---

<sup>4</sup>See the Mollen Commission report investigating police violence in New York and Moe (2011) for the case of teachers unions. Benoit and Dubra (2004) mention the case of a *white wall of silence* in the medical profession emerging from a CNN&TIME investigation available at <http://edition.cnn.com/TRANSCRIPTS/0009/24/impc.00.html>.

of protection for weakest members we must ask if is indeed the case that the demand for labor is typically inelastic. Indeed, the empirical evidence strongly favors inelasticity across industries and countries. For example Lichter, Peichl and Siegloch (2015) do a meta-study of labor demand across industries and the vast bulk of estimates lie above  $-1$ , corresponding to inelastic demand.<sup>5</sup> We then focus on the case in which workers dominate, which is also consistent with the conventional wisdom that there are only a few bad teachers or a few bad cops. In this case, if we consider an alternative information structure and assume the employees can share their information with the firms, then all employees, regardless of type, are *strictly* better off shielding the worst. Hence, not only will labor associations protect shirkers and discourage workers from working too hard, but they will conceal from firms the performance of employees.

Finally, still concentrating on the inelastic case where the optimal labor association plan calls for relatively low overall effort, we shed light on a specific aspect of the trade-off workers face in participating in the association. Specifically, we compare how much utility a worker gets under the optimal association plan versus how much would they get if there is no labor association and the firm observes effort so can provide perfect incentive contracts. In the latter case workers receive a premium because shirkers must be compensated for their effort and this is reflected in the wage all get. When the required compensation is relatively low, however, the premium will be less than the increased wage they will receive if the association plan is in place.

In order to understand under what conditions employee associations can successfully restrict effort of their members in the interest of the group we need a theory of how they provide incentives. We know from the work of Ostrom (1990) and her successors how this can be achieved: groups can self-organize to overcome the free rider problem and provide public goods through peer monitoring and social punishments such as ostracism. Formal theories of this type originate in the work of Kandori (1992) on repeated games with many players and have been specialized to the study of organizations. The basic idea is that

---

<sup>5</sup>In the Appendix we reproduce their figure showing estimates from different studies of the elasticity of labor demand.

groups choose norms consisting of a target behavior for the group members and individual penalties for failing to meet the target; these norms are endogenously chosen in order to advance group interests. Specifically the group designs a mechanism to promote group interests subject to incentive constraints for individual group members, and it provides incentives in the form of punishments for group members who fail to adhere to the norm.<sup>6</sup>

In this paper we build on this theory and show that the optimal target level of average effort in an industry crucially depends on the elasticity of labor demand *and* on the difficulty of monitoring individual behavior. While elasticity of demand determines whether it is optimal to restrain or incentivize effort, monitoring difficulty, which in turn depends on the social network structure of the industry, determines whether it is possible to do so or not. Both elements are therefore necessary for effort quotas to emerge in equilibrium. We show, moreover, that similar considerations apply not only to labor associations but to individual proprietors who sell into the market at a piece rate: we argue that country squires should be “lazy” because they face inelastic output demand and industrialists “energetic” because they face elastic demand.

We are not the first to ask why labor associations protect their weakest members. Our explanation complements existing theories which focus on narrow details of the punishment or production process. Benoit and Dubra (2004) focus on testimony before a court with rules exogenous to the industry and argue that setting up a wall of silence may reduce the probability of incurring in a type II error. Muehlheusser and Roider (2008) focus on team production and emphasize the need for cooperation in such a setting.

We should also indicate that most occupations do have a limited form of performance based pay in the form of a probationary period during which or at the end of which the employee can be laid off without cost to the firm. Our focus is rather on post probationary incentives: naturally during probation we would expect shirkers to conceal their type so that those laid off would most likely be types incapable of providing effort.

---

<sup>6</sup>See for example Levine and Modica (2016) and Dutta, Levine and Modica (2021).

## 2. The Model

Let  $x$  denote labor input and suppose that the value of that input to the industry  $U(x)$  is strictly differentially concave up to a satiation level  $X$ , so that the marginal product of labor  $U'(x)$  is positive and declining with input for  $x < X$ . We assume moreover that the revenue function  $R(x) = xU'(x)$  is concave, that is, marginal revenue is declining with input. Firms are competitive and simply convert labor to output.

We denote by  $n$  the size of the labor force in the industry and by  $e^i \in [0, 1]$  the effort provided by employee  $i$ , with average employee effort denoted by  $\bar{e}$ . Total labor input is thus  $x = n\bar{e}$ . The opportunity cost of the  $i$ th employee is  $v(i)$ , which we assume to be strictly differentially increasing with  $v'(i) > 0$ .

There are two types of employees, *workers*  $w$  and *shirkers*  $s$ . Types are realized *ex-post* after employment and they are private information. The exogenous probability of an employee being of type  $w$  is  $\gamma$ . Letting  $W$  denote the employees' wage, when  $n$  employees are working the net utility of the  $i$ th employee working is  $W - v(i) - \mu^s e^i$  for a shirker and  $W - v(i) - \mu^w (1 - e^i)$  for a worker, where the last term is the cost of effort, with  $\mu^s$  and  $\mu^w$  both positive numbers. This captures our simple assumption that workers prefer to provide effort because they get satisfaction out of a job well done, while for shirkers effort is costly.

Employees collectively face a mechanism design problem: they can set an effort quota and observe a noisy signal of whether the quota is adhered to. We consider the two alternative cases of a minimum and a maximum quota on effort. Denote by  $\phi^-$  a minimum quota on effort meaning that only effort levels  $e \in [\phi^-, 1]$  are acceptable, and by  $\phi^+$  a maximum quota meaning that only effort levels  $e \in [0, \phi^+]$  are acceptable. While individual efforts are not observable, each employee - conditional on her effort choice - produces a public signal  $z^i \in \{0, 1\}$  where 1 means good, adhered to the effort quota and 0 means bad, violated the quota. Furthermore, if the quota was adhered to, the probability of producing a bad signal is  $\pi$  while if it is violated it is  $\pi' > \pi$ . If an employee is seen to have a bad signal, an endogenous utility punishment  $P$  is issued.

A feasible employee mechanism is an incentive compatible choice of  $\phi$  and

$P$ . Average effort for an incentive compatible minimum quota is  $\bar{e} = (1-\gamma)\phi^- + \gamma$  (workers provide full effort and shirkers make the minimum effort allowed) and with an incentive compatible maximum quota is  $\bar{e} = \gamma\phi^+$  (only workers provide effort, the maximum allowed). The cost of effort to the group has two components, a direct cost and a monitoring cost. The average direct cost for a minimum quota is  $D = (1-\gamma)\mu^s\phi^-$  and for a maximum quota  $D = \gamma\mu^w(1-\phi^+)$ . The average monitoring cost for a quota is  $M = \pi P$ , that is the expected cost of punishment “on the equilibrium path” when everyone adheres to the quota. The total average cost of effort is therefore  $C = D + M$ . Under “no quota”  $\phi^- = 0$  or  $\phi^+ = 1$  average effort is  $\gamma$  and  $C = 0$ .

Following the choice of mechanism by employees, market clearing takes place. Initially we assume that firms cannot observe any signal of employee effort so must pay a fixed wage  $W$  to all employees. On the demand side we assume competitive firms which make zero profits; since profit is  $n\bar{e}U'(n\bar{e}) - Wn = n(\bar{e}U'(n\bar{e}) - W)$  we have  $W = \bar{e}U'(n\bar{e})$ . On the supply side  $W - C \leq v(n)$  with equality if  $n > 0$ . In other words market clearing is given by  $\bar{e}U'(n\bar{e}) - C \leq v(n)$  with equality if  $n > 0$ . Because  $U(x)$  is strictly concave and  $v(n)$  is strictly increasing this has a unique solution, denoted by  $\hat{n}(\bar{e}, C)$ . Observing that from the supply side higher  $n$  means both more employees and (since the wage is increasing in  $v$ ) that all employees receive a higher utility, the unambiguous objective of the employees, regardless of type, is to maximize  $\hat{n}$ . We assume that  $\gamma U'(0) > v(0)$  so that  $\hat{n}(\gamma, 0) > 0$ . A useful benchmark will be the  $x^\gamma$  uniquely defined under these assumptions by  $\gamma U'(x^\gamma) = v(x^\gamma/\gamma)$  which is the market clearing output when workers work and shirkers shirk.

### 3. Total Cost of Effort

How much does it cost to optimally implement a target level of average effort  $\bar{e}$ ? Intuitively, it should depend on the ex-ante likelihood  $\gamma$  of an employee being of the  $w$  type. To implement a target level  $\bar{e} > \gamma$ , incentives for shirkers to provide effort must be used. Otherwise, to implement  $\bar{e} < \gamma$ , incentives for workers not to provide effort must be used. Defining monitoring difficulty as  $\theta = \pi/(\pi' - \pi)$  we have the following result.

**Theorem 1.** *The cost of implementing a target level of average effort  $\bar{e}$  is given by*

$$C(\bar{e}) = \begin{cases} -(\mu^w(\gamma + \theta)/\gamma)(\bar{e} - \gamma) \equiv -c^w(\bar{e} - \gamma) & \bar{e} < \gamma \quad (\text{low}) \\ 0 & \bar{e} = \gamma \quad (\text{intermediate}) \\ (\mu^s(1 - \gamma + \theta)/(1 - \gamma))(\bar{e} - \gamma) \equiv c^s(\bar{e} - \gamma) & \bar{e} > \gamma \quad (\text{high}) \end{cases}$$

*Proof.* Incentive compatibility for a minimum quota is that shirkers must prefer providing  $\phi^-$  to not providing effort. This is  $-\mu^s\phi^- - \pi P \geq -\pi'P$ . The optimal mechanism must minimize  $C$  hence  $P$ , therefore this constraint must hold with equality. This gives  $M = \theta\mu^s\phi^-$  and  $C = \mu^s(1 - \gamma + \theta)\phi^-$ . For the maximum quota we have for workers  $-\mu^w(1 - \phi^+) - \pi P \geq -\pi'P$ . This gives  $M = \theta\mu^w(1 - \phi^+)$  and  $C = \mu^w(\gamma + \theta)(1 - \phi^+)$ . For  $\bar{e} = \gamma$  no incentives are needed,  $P = 0$  and maximal effort by workers  $e = 1$  and minimal effort by shirkers  $e = 0$  are incentive compatible and have associated cost  $C = 0$ . If  $\bar{e} > \gamma$ , a minimum quota must be established so that  $\bar{e} = (1 - \gamma)\phi^- + \gamma$ , with corresponding cost  $C = \mu^s(1 - \gamma + \theta)\phi^- = \mu^s(1 - \gamma + \theta)(\bar{e} - \gamma)/(1 - \gamma)$ . If  $\bar{e} < \gamma$ , a maximum quota must be established so that  $\bar{e} = \gamma\phi^+$ , with corresponding cost  $C = \mu^w(\gamma + \theta)(1 - \phi^+) = \mu^w(\gamma + \theta)(1 - \bar{e}/\gamma) = \mu^w(\gamma + \theta)(\gamma - \bar{e})/\gamma$ .  $\square$

#### 4. The Optimal Mechanism

Inverting  $W = \bar{e}U'(n\bar{e})$  gives demand  $x = n\bar{e} = (U')^{-1}(W/\bar{e})$  with derivative  $dx/dW = 1/[U''(x)\bar{e}]$ , so the (non-positive) elasticity of labor demand with respect to the wage is  $\eta(x) \equiv dx/dW \cdot W/x = U'(x)/(xU''(x))$ . Slightly abusing notation we also let  $\hat{n}(\bar{e}) \equiv \hat{n}(\bar{e}, C(\bar{e}))$ . An optimal target level of average effort is a choice  $\bar{e} = \hat{e}$  which maximizes  $\hat{n}(\bar{e})$ . We let  $\hat{x} = \hat{n}(\hat{e})\hat{e}$ . In particular our benchmark  $x^\gamma$  market clearing output when workers work and shirkers shirk is given by  $x^\gamma = \hat{n}(\gamma)\gamma$ . Recall that  $R(x) = xU'(x)$  is the (concave) revenue function. We are now ready to state our first main result.

**Theorem 2.** *The optimal target level of average effort  $\hat{e}$  is unique and*

- (low) *If  $R'(x^\gamma) < -c^w$  then  $\hat{e} < \gamma$  and  $\eta(\hat{x}) > -1$  that is a maximum quota is optimal and demand is inelastic;*

- (intermediate) If  $-c^w \leq R'(x^\gamma) \leq c^s$  then  $\hat{e} = \gamma$  that is no quota is optimal;
- (high) If  $R'(x^\gamma) > c^s$  then  $\hat{e} > \gamma$  and  $\eta(\hat{x}) < -1$  that is a minimum quota is optimal and demand is elastic.

In the low case  $\hat{e}$  is uniquely determined as the solution of  $R'(\hat{n}\hat{e}) + c^w = 0$  and in the high case by  $R'(\hat{n}\hat{e}) - c^s \geq 0$  with equality if  $\hat{e} < 1$ .

Before proving this central result of the paper two observations are worth mentioning. First, since  $c^w$  and  $c^s$  increase in  $\theta$ , for large enough  $\theta$  - that is an inefficient monitoring technology - we are in the intermediate domain and therefore setting a quota is definitely not optimal. This is fairly intuitive since if monitoring is very difficult, the cost of implementing any level of effort different from the *natural* level  $\gamma$  is extremely costly. Second, notice that if  $\eta(\hat{x}) \geq -1$  then we must be in the intermediate or low domain. That is, if labor demand is inelastic, it is never optimal to incentivize shirkers and if incentives are provided they are for workers not to work too hard. On the other hand, if demand is elastic all employees will agree that workers should not be discouraged.

*Proof.* By Theorem 1 since  $C(\gamma) = 0$  and the assumption that  $\gamma U'(0) > v(0)$  we have  $\hat{n}(\gamma) \equiv \hat{n}(\gamma, 0) > 0$  hence  $\hat{n}(\hat{e}) \geq \hat{n}(\gamma) > 0$ . When  $\hat{n}(\bar{e}) > 0$  we must have  $\bar{e}U'(\hat{n}(\bar{e})\bar{e}) - C(\bar{e}) - v(\hat{n}(\bar{e})) = 0$  and since  $v$  is positive, this requires  $U'(\hat{n}(\bar{e})\bar{e}) > 0$ ; this implies that  $\hat{n}(\hat{e})\hat{e} < X$  ( $X$  being the satiation level of utility). Consider two domains: in the higher domain  $\bar{e} \geq \gamma$  and  $\hat{n}\bar{e} \leq X$ ; in the lower domain  $\bar{e} \leq \gamma$  and  $\hat{n}\bar{e} \leq X$ . Then  $h(\bar{e}, \hat{n}) \equiv \bar{e}U'(\hat{n}\bar{e}) - C(\bar{e}) - v(\hat{n}) = 0$  is smooth in each of these domains and in either one by using the implicit function theorem we obtain

$$\frac{d\hat{n}}{d\bar{e}} = -\frac{U'(\hat{n}\bar{e}) + \hat{n}\bar{e}U''(\hat{n}\bar{e}) - C'(\bar{e})}{\bar{e}^2U''(\hat{n}\bar{e}) - v'(\hat{n})} = -\frac{R'(\hat{n}\bar{e}) - C'(\bar{e})}{\bar{e}^2U''(\hat{n}\bar{e}) - v'(\hat{n})}.$$

Since the denominator  $\bar{e}^2U''(\hat{n}\bar{e}) - v'(\hat{n}) < 0$ , at an interior local maximum of  $\hat{n}(\bar{e})$  where  $\bar{e} \neq \gamma$  it must be that  $R'(\hat{n}\bar{e}) - C'(\bar{e}) = 0$ . Computing the second derivative where  $R'(\hat{n}\bar{e}) - C'(\bar{e}) = 0$  and  $\bar{e} \neq \gamma$  yields

$$\frac{d^2\hat{n}}{d^2\bar{e}} = -\frac{\hat{n}R''(\hat{n}\bar{e})}{\bar{e}^2U''(\hat{n}\bar{e}) - v'(\hat{n})} < 0,$$



where we used that  $C''(\bar{e}) = 0$  from Theorem 1. This implies that  $R'(\hat{n}\bar{e}) - C'(\bar{e}) = 0$  is always a local maximum and not a local minimum. From Theorem 1, we know that in the lower domain  $R'(\hat{n}\bar{e}) - C'(\bar{e}) = R'(\hat{n}\bar{e}) + c^w$  while in the higher domain  $R'(\hat{n}\bar{e}) - C'(\bar{e}) = R'(\hat{n}\bar{e}) - c^s$ . Hence in the lower domain if  $R'(x^\gamma) + c^w \geq 0$  there can be no local maximum with  $\bar{e} < \gamma$ , while in the higher domain if  $R'(x^\gamma) - c^s \leq 0$  there can be no local maximum with  $\bar{e} > \gamma$ . Hence if both these conditions hold we are in the intermediate case. Moreover, if the first condition fails we must have  $R'(x^\gamma) < 0$  while if the second fails we must have  $R'(x^\gamma) > 0$  so at most one of them fails. If the first fails - that is  $R'(x^\gamma) + c^w < 0$  - then there must be a unique local maximum in the strict lower domain which, since the second condition holds (that is,  $R'(x^\gamma) - c^s < 0$ ), is a global maximum. Similarly if the second fails - so that  $R'(x^\gamma) - c^s > 0$  - there must be a unique global maximum in the strict higher domain. In the lower domain the first order condition  $R'(\hat{n}\hat{e}) - c^s = 0$  uniquely determines the maximum, while in the higher domain the constraint  $\hat{e} \leq 1$  may bind, so the condition is that given in the theorem. Finally, since  $R'(x^\gamma) > c^s$  implies the higher domain, from the first order condition  $R'(\hat{n}\hat{e}) - c^s \geq 0$  we see that  $R'(\hat{n}\hat{e}) \geq c^s > 0$ , and similarly since  $R'(x^\gamma) < -c^w$  implies the lower domain, from the first order condition  $R'(\hat{n}\hat{e}) + c^w = 0$  we see that  $R'(\hat{n}\hat{e}) = -c^w < 0$ . Since with  $R''(x) < 0$  we have  $\eta(x) < -1$  if and only if  $R'(x) > 0$  and  $\eta(x) > -1$  if and only if  $R'(x) < 0$  the results about demand elasticity follows.  $\square$

### *Interpreting inequalities and a parametric example*

Given monitoring difficulty and the  $\mu$  coefficients  $\mu^w$  and  $\mu^s$ , the inequalities involving  $R'(x^\gamma)$  depend on the relative size of workers and shirkers. Precisely, the condition  $R'(x^\gamma) < -c^w$  tends to hold when  $\gamma$  is large, while  $R'(x^\gamma) > c^s$  is more likely to hold for small  $\gamma$ . The reason is that  $c^w$  decreases in  $\gamma$  while  $c^s$  increases in  $\gamma$ ; on the other hand  $x^\gamma$  increases with  $\gamma$ , so  $R'(x^\gamma)$  decreases. To verify this:  $\hat{n}(\gamma)$  is defined by  $\gamma U'(\hat{n}\gamma) = v(\hat{n})$ , so  $\hat{n}'(\gamma) = \gamma^2 / (v'(n) - U''(n\gamma)) > 0$  whence  $dx^\gamma/d\gamma = d(\hat{n}(\gamma)\gamma)/d\gamma > 0$ . In conclusion large  $\gamma$  favors the inequality  $R'(x^\gamma) < -c^w$  while small  $\gamma$  favors  $R'(x^\gamma) > c^s$ .

Consider specifically the quadratic  $U(x) = x(a - x/a)$  and linear  $v(i) = v \cdot i$ . Also take  $\theta = \mu^s = \mu^w = 1$ . In the Appendix we show that  $R'(x^\gamma) < -c^w$  for

$\gamma$  close enough to 1 and  $va^2 - 2(1-v)a + 4 < 0$  (for example:  $v = 0.1$  and  $2.6 < a < 15$ ) and  $R'(x^\gamma) > c^s$  for  $\gamma$  close enough to 0 and  $a > 2$ .

## 5. The Blue Wall of Silence

Suppose now that in addition to choosing a quota and a punishment the labor association can also choose whether or not to voluntarily reveal the realization of the individual signals  $z^i$  to firms. If they do so this enables firms to set wages (after the mechanism  $\phi, P$  is chosen) conditional on the value of the signal, where  $W(0)$  is conditional on bad signal and  $W(1)$  is conditional on good signal.<sup>7</sup> Recall that if the norm is followed the bad signal is generated with probability  $\pi$ . Knowing the mechanism put in place by the labor association firms choose these wages optimally. We need one technical proviso in this. When there is a maximum quota, firms may wish to induce shirkers to produce to that quota. Workers can violate the quota and generate the bad signal with the high probability  $\pi'$  by exceeding the quota by any positive  $\epsilon$ . If this is to their advantage they would always prefer to produce slightly less while still violating the quota leading to a trivial non-existence problem. Hence following Simon and Zame (1990) we introduce an endogenous tie-breaking rule, and assert that if a type can violate a quota by  $\epsilon$  then, if they exactly meet the quota, they can choose between  $\pi$  and  $\pi'$ , that is they can choose if they wish the probability of generating a bad signal. Given this technical condition we are ready to state our second main result.

**Theorem 3.** *Consider the incentives to reveal or not the realization of the signals  $z^i$  to firms:*

*(Inelastic demand) If  $R'(x^\gamma) \leq 0$  the labor association weakly prefers not to reveal the signal and if  $R'(x^\gamma) < -c^w$  it strictly prefers not to.*

*(Elastic demand) If  $R'(x^\gamma) \geq 0$  the labor association weakly prefers to reveal the signal and if  $R'(x^\gamma) > c^s$  it strictly prefers to.*

The intuition is this. If demand is elastic, the labor association never wants to discourage workers and sometimes wants to incentivize shirkers by setting

---

<sup>7</sup>We assume that negative wages are not feasible.

a minimum quota. By revealing individual signals to the firm, the labor association can reduce the monitoring cost of implementing a minimum quota since the firm can punish deviations by setting the payment  $W(0)$  to a bad signal very low. In a sense the labor association, by revealing the signals, can “outsource” part of the cost of punishing to the firm. On the other hand, if demand is inelastic, the labor association never wants to incentivize shirkers and sometimes wants to discourage workers by setting a maximum quota. By revealing the signal to the firm, the labor association increases the monitoring cost of implementing a maximum quota since the firm can reward deviations of workers setting a relatively high payment to a bad signal  $W(0)$ . In a sense the labor association, by revealing the signals is increasing its cost of restraining workers from exerting effort.

Note that as a matter of practice the labor association may effectively suppress the signal by requiring firms to set non-contingent wages as is often done in union contracts where wages must be based only on seniority and not signals of job performance.

*Proof.* For a given market wage  $W$  the labor association must choose a  $\phi, P$  such that there exists a firm optimal choice  $\pi W(0) + (1 - \pi)W(1) = W$  that is incentive compatible for both types. Define the premium for a good signal as  $\Delta = W(1) - W(0)$ . In the case of a minimum quota the binding constraint is that shirkers must prefer providing  $\phi^-$  to not providing effort. This is

$$\mu^s \phi^- \leq (\pi' - \pi)(P + \Delta).$$

For the maximum quota we have for workers

$$\mu^w(1 - \phi^+) \leq (\pi' - \pi)(P + \Delta).$$

Observe that a deviation from  $\Delta$  to  $\Delta'$  by the firm only impacts its profits if it is chosen so that the constraint is violated: this means  $\Delta' < \Delta$ .

Consider first the elastic case. Let  $\hat{\phi}^-, \hat{P}, \hat{W}$  be the optimum without revealing the signal. If this is a no quota then the firm cannot provide incentives either, so revealing makes no difference and the labor association weakly prefers

to reveal. If there is a quota then consider setting  $P = \mu^s \hat{\phi}^- / (\pi' - \pi) - W / (1 - \pi)$  - recall that the punishment value in the original model is  $\mu^s \hat{\phi}^- / (\pi' - \pi)$ . If the firm chooses  $\Delta = W(1) = W / (1 - \pi)$  and  $W(0) = 0$  this is optimal for the firm since  $\Delta$  can be increased while still paying  $W$  only by raising  $W(1)$  which raises firm costs without changing worker behavior; and decreasing  $\Delta$  will induce the workers to violate the constraint decreasing output, but since the firm must still incur an expected cost of  $W$  this lowers profits. Hence the labor association can obtain the same result in terms of effort and wage by revealing while incurring strictly less punishment cost, so it strictly prefers to reveal.

Consider then the inelastic case. If there is revelation it is still the case that no quota is at least as good as a minimum quota. In fact, while no quota has no cost, a minimum quota can only raise output over no quota and raising output in an incentive compatible way has a non-negative punishment cost. If the optimum with revelation is no quota then either that was the optimum without revelation in which case no revelation is weakly preferred, or it was not in which case no revelation is strictly preferred.

Suppose then that the equilibrium  $\hat{\phi}^+, \hat{P}, \hat{W}, \hat{\Delta}$  with revelation has a maximum constraint. Recall that in the original model the punishment value is  $\mu^w(1 - \hat{\phi}^+) / (\pi' - \pi)$ . Suppose that

$$\mu^w(1 - \hat{\phi}^+) \geq (\pi' - \pi)\hat{P}.$$

Consider  $\Delta < 0$ , then

$$\mu^w(1 - \hat{\phi}^+) > (\pi' - \pi)(\hat{P} + \Delta)$$

violating the constraint and inducing the workers to work. Per employee this costs the firm  $(\pi' - \pi)(W(0) - W) = -(\pi' - \pi)(1 - \pi)\Delta$  and the gain is  $U'(n\bar{e})\gamma(1 - \hat{\phi}^+)$ . In other words, by choosing  $\Delta$  close to zero the firm could increase its profit. Hence if the maximum constraint is to be incentive compatible for the firm it must be that

$$\mu^w(1 - \hat{\phi}^+) < (\pi' - \pi)\hat{P}.$$

In this case the labor association can get the same outcome by not revealing and choosing  $P = \mu^w(1 - \hat{\phi}^+)/(\pi' - \pi)$  strictly reducing cost and is therefore better off.  $\square$

## 6. Utility of Workers

To what extent are workers content with their colleagues shirking? This is a tricky question to ask in the current context because we assume that *ex ante* employees do not know their own type. If they did we would need to consider the possibility that firms would introduce screening contracts in an effort to lure workers rather than shirkers. Never-the-less we can get an idea of this with the following conceptual experiment. Specifically, we can ask how much utility a worker gets under the optimal labor association plan versus how much would they get if there is no labor association and the firm observes effort so can provide perfect incentive contracts.

For simplicity we will focus on the case in which  $-c^w \leq R'(x^\gamma) < 0$ . We assume as well that  $\mu^s < U'(n)$  so that it is efficient for shirkers to provide full effort. The marginal revenue condition says that optimal quota is no quota and that since marginal revenue is negative the association would benefit from lower effort. On the other hand  $c^w$  is sufficiently large so from Theorem 2 enforcement is too costly and the association does not try to restrict the output of workers.

**Theorem 4.** *Suppose that  $-c^w = -\mu^w(\gamma + \theta)/\gamma \leq R'(x^\gamma) < 0$ . Then there exists a  $\bar{\mu}^s > 0$  such that for  $0 < \mu^s < \bar{\mu}^s$  workers are strictly better off with a labor association.*

*Proof.* Suppose there is no labor association and the firm can observe effort. Since shirkers provide full effort then the equilibrium wage is  $W^{NA} = U'(n^{NA}) = v(n^{NA}) + \mu^s(1 - \gamma)$  where  $n^{NA}$  is equilibrium employment. This is since shirkers - fraction  $1 - \gamma$  of the population - must be compensated for their effort. With a labor association, since  $\bar{e} = \gamma$  and  $C = 0$ , the wage is given by  $W^A = \gamma U'(\gamma n^A) = v(n^A)$ . Let  $n^C$  be defined by  $W^C = U'(n^C) = v(n^C)$ . Because  $R' < 0$  it must be that  $W^A > W^C$ . As  $W^C$  corresponds to  $\mu^s = 0$  by continuity for small enough  $\mu^s$  we have  $W^A > W^{NA}$ .  $\square$

The point is that without the labor association if the firm observes effort then workers receive a premium because shirkers must be compensated for their effort. However, if shirkers do not require much compensation this premium will be less than the increased wage they will receive if instead shirkers do not provide effort.

## 7. Piece-rate Payments

We do not mean to pick on workers as being especially lazy as compared to, for example, proprietors. Proprietors unlike workers cannot contract to be paid regardless of effort - they (as do some workers such as garment workers) are paid a piece-rate proportional to effort. None-the-less similar considerations of elasticity and lack of effort apply. We turn here to proprietors who are paid a piece rate and for simplicity take the neutral assumption that all are identical and that effort has neither cost nor benefit.

We consider a fixed force  $N$  of identical proprietors who costlessly provide effort  $e^i \in [0, 1]$ . As before if average effort is  $\bar{e}$  total output is  $x = N\bar{e}$ . We continue to assume the value of output  $U(x)$  is strictly differentially concave up to a satiation level  $X > N\bar{e}$  and that the revenue function  $R(x) = xU'(x)$  is concave. Now, however, proprietors face a constant marginal cost  $c$  of other inputs used in producing output and are paid individually for the output they produce, so that the profit of a proprietor is  $U'(x)e - ce$ . We assume that  $U'(N) > c$  so that the market clearing effort level absent any incentives is  $N$  with corresponding market price  $U'(N)$ . In this context  $\eta(x) = -U'(x)/xU''(x)$  is the elasticity of demand for output.

The group of proprietors also faces a mechanism design problem: they can set an effort quota and observe a noisy signal of whether the quota is adhered to by a member. As before there can be either a minimum quota  $\phi^-$  or a maximum quota  $\phi^+$ . Although the market implicitly measures the effort of each individual proprietor we assume that this information is not so easy for other proprietors to observe. Hence we continue to assume that individual efforts are not observable so that other proprietors observe only a noisy public signal  $z^i \in \{0, 1\}$  of adherence to the quota where again 1 is good and 0 is

bad. The probabilities of the signal remain  $\pi' > \pi$  as the quota is not or is adhered to. Proprietors can impose costly social punishments on each other of  $P$ . Roughly speaking we assume that proprietors, whether butchers, country squires, or industrialists like to socialize with people in the same line of business so that ostracism from the association of proprietors is costly.

In formulating a precise result it will be convenient to work with the *inverse elasticity of the price cost margin*

$$\eta(x) \equiv -\frac{U'(x) - c}{xU''(x)}.$$

The monopoly solution is at  $R'(x^m) = c$  that is  $\eta(x^m) = 1$ , while the competitive solution  $x^c$  has  $U'(x^c) = c$  so  $\eta(x^c) = 0$ ; also observe that  $\eta(0) = \infty$ . In place of assuming that marginal revenue is declining with output we will use here the obvious regularity condition that  $\eta(x)$  is declining. When  $c = 0$  we have  $\eta(x)$  is simply the elasticity of demand, and this is the usual assumption that demand elasticity is declining with output. We can now state our main result in the case of piece-rate payments.

**Theorem 5.** *A minimum quota is never used. A binding maximum quota is used if and only if*

$$\eta(N) < \frac{\pi' - \pi}{\pi'}.$$

*Proof.* Individual proprietor profit is given by  $U'(N\bar{e})e - ce$ . Since  $\bar{e} \leq 1$ ,  $U'(N) > c$  and decreasing, it follows that proprietors would like to increase effort over any quota so minimum quotas would be useless.

The maximum incentive constraint is  $U'(N\phi^+)\phi^+ - c\phi^+ - \pi P \geq U'(N\phi^+) - c - \pi'P$  or

$$P = \frac{(U'(N\phi^+) - c)(1 - \phi^+)}{\pi' - \pi}.$$

Hence profits of a member of the association is given by

$$\begin{aligned} & U'(N\phi^+)\phi^+ - c\phi^+ - \frac{\pi}{\pi' - \pi} (U'(N\phi^+) - c)(1 - \phi^+) \\ &= (U'(N\phi^+) - c) \left[ \frac{\pi'}{\pi' - \pi} \phi^+ - \frac{\pi}{\pi' - \pi} \right] \end{aligned}$$

Differentiate with respect to  $\phi^+$  to get

$$\frac{\pi'}{\pi' - \pi} [N\phi^+U''(N\phi^+) + U'(N\phi^+) - c] - \frac{\pi}{\pi' - \pi} NU''(N\phi^+).$$

This has the same sign as

$$-\frac{U'(N\phi^+) - c}{N\phi^+U''(N\phi^+)} - \left[1 - \frac{\pi}{\pi'} \frac{1}{\phi^+}\right] = \eta(N\phi^+) - \left[1 - \frac{\pi}{\pi'} \frac{1}{\phi^+}\right].$$

As this is decreasing in  $\phi^+$  we see that the condition for a binding maximum constraint is indeed

$$\eta(N) - \left[1 - \frac{\pi}{\pi'}\right] < 0.$$

□

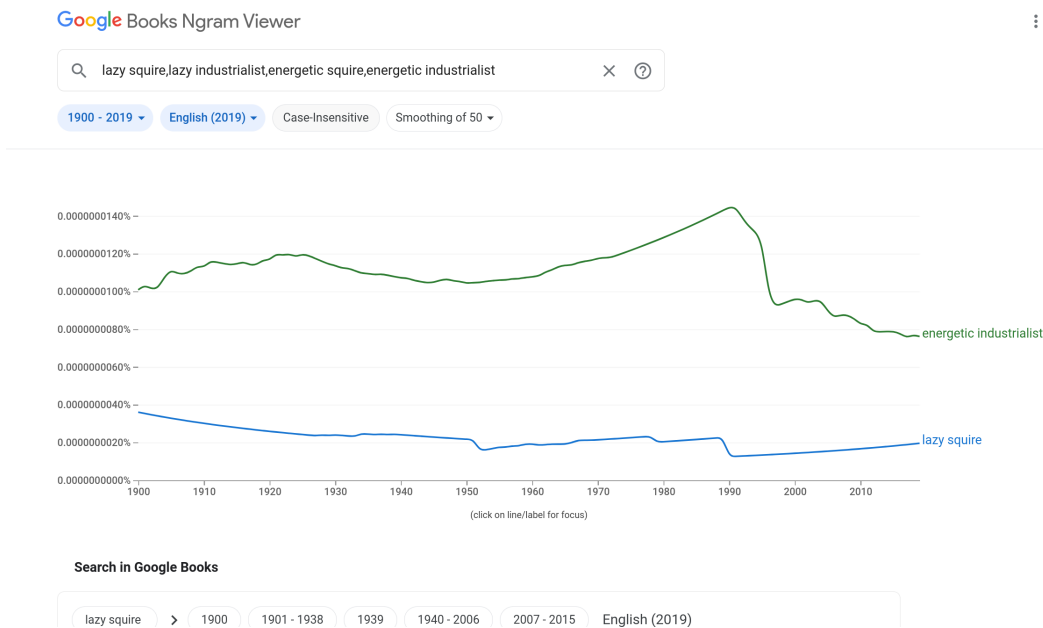
### *Squires versus Industrialists*

Are there cases where proprietors use social incentive to restrict effort? We argue that this was exactly what the British land-owning nobility - the “country squires” engaged in during the 18th and 19th Centuries. The country squire is infamous in British literature for their drunken lazy ways and their engagement in social activities such as throwing parties and fox hunting: Fielding (1742) is scathing in his description of the country squire. The Sicilian aristocracy is equally well known for the same kind of lifestyle (and in fact their British peers were not infrequently among the guests at their lavish parties).

The country squires produced mostly staple agricultural products, mostly grain and primarily for domestic consumption; and demand for these products is known to be inelastic - see, for example, Andreyeva, Long, and Brownell (2010). Since inelastic demand implies an inelastic inverse elasticity of the price cost margin, Theorem 5 implies that a social norm of “spend all your time having parties and fox-hunts rather than running your farm” makes sense - and has relatively low monitoring costs since it is easy to see if your colleagues are inviting you to lavish parties and fox-hunts. In this view, then, the “laziness” of country squires was simply a rational way to restrict output and exercise monopoly power.



In contrast to country squires industrialists were not famed for their laziness. Our Ngram reported below examines the 20th Century English language literature for lazy squire, lazy industrialist, energetic squire, and energetic industrialist. As can be seen squires are frequently described as lazy and industrialists as energetic, but pretty much never the other way around. If indeed demand for industrial products is sufficiently elastic our Theorem 5 makes sense of this.



Intuitively, manufacturers exporting goods face fairly elastic demand due to the presence of many substitutes. From Stokey (2001) we find that indeed during the early industrial revolution output and revenue increased hugely, indicating a high elasticity. Specifically Stokey (2001) reports that from 1780 to 1850 GDP grew by a factor of 3.65 and industrial output by a factor of 6.07 so that there was a large increase in the relative share of industrial output. On the other hand capital's share of GDP rose from .35 to .44. A large relative increase in output share with an increased profit share indicates that indeed demand must have been highly elastic.

## 8. Conclusion

Discouraging workers from working and imposing a blue wall of silence is inefficient, and of course particularly harmful to consumers. What can be done about it?

One policy discussed recently is to “defund the police.” We take this to mean conditioning wages on some sort of measure of average performance: if the police force as a whole fails to live up to some standard they are all fired or their wages are reduced. As in our model, police forces are relatively competitive: even within a single jurisdiction there are typically many police forces, and of course different suburbs often have their own police forces. In the USA as a whole there are roughly 18,000 different police forces.<sup>8</sup> While in reality unlike the model it is not costless for police to get a job with another force, it is never-the-less hard to see how such a threat of collective punishment by a single jurisdiction would have much effect on the behavior of the labor association.

A second alternative is to improve transparency so that performance of individual police officers is observed more easily and to allow performance based pay - that is, to attack the blue wall of silence. Roughly speaking this corresponds to forcing the labor association to reveal its signal and while our Theorem 3 shows that the labor associations will not much like this, it also shows that it weakens their ability to discourage effort. Reforms in this direction are consequently more likely to be successful.

---

<sup>8</sup>[https://cops.usdoj.gov/pdf/taskforce/taskforce\\_finalreport.pdf](https://cops.usdoj.gov/pdf/taskforce/taskforce_finalreport.pdf)

## References

- Andreyeva, T., M. W. Long and K. D. Brownell (2010): "The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food," *American Journal of Public Health* 100: 216-222.
- Benoît, Jean-Pierre, and Juan Dubra (2004). "Why do good cops defend bad cops?." *International Economic Review* 45: 787-809.
- Chin, G. J., and S. C. Wells, S. C. (1997): "The blue wall of silence as evidence of bias and motive to lie: A new approach to police perjury," *U. Pitt. L. Rev.* 59, 233.
- Dutta, R., D. K. Levine and S. Modica (2021): "The Whip and the Bible: Punishment Versus Internalization," forthcoming *Journal of Public Economic Theory*.
- Fielding, Henry (1742): *Joseph Andrews*.
- Iris, Mark (1998): "Police Discipline in Chicago: Arbitration or Arbitrary," *J. Crim. L. & Criminology* 89: 215.
- Kandori, M. (1992): "Social Norms and Community Enforcement," *The Review of Economic Studies* 59: 63-80.
- Khaliq, Amir A., Hani Dimassi, Chiung-Yu Huang, Lutchnie Narine, and Raymond A. Smego Jr. (2005): "Disciplinary action against physicians: who is likely to get disciplined?: *The American journal of medicine* 118: 773-777. (5% of complaints from physicians)
- Levine, D. K., A. Mattozzi and S. Modica (2020): "Trade Associations: Why Not Cartels?" *International Economic Review*, forthcoming
- Levine, D. K. and S. Modica (2016): "Peer Discipline and Incentives within Groups," *Journal of Economic Behavior and Organization* 123: 19-30
- Levine, D. K. and W. Pesendorfer (1995): "When are Agents Negligible?" *American Economic Review*.
- Levine, D. K. and S. Modica (2017): "Size, Fungibility, and the Strength of Lobbying Organizations", *European Journal of Political Economy* 49: 71-83
- Lichter, A., A. Peichl, A. and S. Siegloch. (2015): "The own-wage elasticity of labor demand: A meta-regression analysis," *European Economic Review* 80: 94-119.
- Moe, T. (2011): *Special Interest: Teachers Unions and America's Public Schools*, Brookings Institution Press.
- Muehlheusser, Gerd, and Andreas Roider (2008): "Black sheep and walls of silence," *Journal of Economic Behavior & Organization* 65: 387-408.
- Ostrom, E. (1990): *Governing the commons: the evolution of institutions for collective action*, Cambridge university press.

- Simon, L. K. and W. R. Zame (1990): “Discontinuous Games and Endogenous Sharing Rules,” *Econometrica* 58: 861-872.
- Stokey, Nancy L. (2001): “A quantitative model of the British industrial revolution, 1780–1850,” *Carnegie-Rochester conference series on public policy* 55: 55-109.
- Zieger, R. H. (1994): *American workers, American unions*, Johns Hopkins Univ. Press.

## Appendix

A *parametric example*:

We analyze the quadratic  $U(x) = x(a - x/a)$  and linear  $v(i) = v \cdot i$ . Also take  $\theta = \mu^s = \mu^w = 1$ . We show that  $R'(x^\gamma) < -c^w$  for  $\gamma$  close enough to 1 and  $va^2 - 2(1 - v)a + 4 < 0$  (for example:  $v = 0.1$  and  $2.6 < a < 15$ ) and  $R'(x^\gamma) > c^s$  for  $\gamma$  close enough to 0 and  $a > 2$ .

Observe that  $U' = a - 2x/a$ ,  $R(x) = x(a - 2x/a) = ax - 2x^2/a$  and  $R'(x) = a - 4x/a$  (which is negative for  $x > a^2/4$ ) and that

$$c^w = 1 + \frac{1}{\gamma}, \quad c^s = 1 + \frac{1}{1 - \gamma}$$

The equality  $\bar{e}U'(n\bar{e}) - C = v(n)$  reads

$$\begin{aligned} \bar{e}(a - 2n\bar{e}/a) - C &= vn \\ a\bar{e} - 2n\bar{e}^2/a - C &= vn \\ \hat{n}(\bar{e}, C) &= \frac{a\bar{e} - C}{v + 2\bar{e}^2/a} \end{aligned}$$

and in particular

$$\hat{n}(\gamma) = \frac{\gamma a}{v + 2\gamma^2/a} \quad \text{and} \quad x^\gamma = \hat{n}(\gamma)\gamma = \frac{a\gamma^2}{v + 2\gamma^2/a}$$

so

$$\begin{aligned} R'(x^\gamma) &= a - \frac{4}{a} \frac{a\gamma^2}{v + 2\gamma^2/a} = a - \frac{4\gamma^2}{v + 2\gamma^2/a} \\ &= \frac{a(v + 2\gamma^2/a) - 4\gamma^2}{v + 2\gamma^2/a} = \frac{av - 2\gamma^2}{v + 2\gamma^2/a} \end{aligned}$$

Next we show that  $R'(x^\gamma) < -c^w$  for  $\gamma$  close enough to 1. For  $\gamma = 1$  we

want

$$\frac{av - 2}{v + 2/a} < -2$$

$$av - 2 + 2v + 4/a < 0$$

$$av - 2 + 2v + 4/a < 0$$

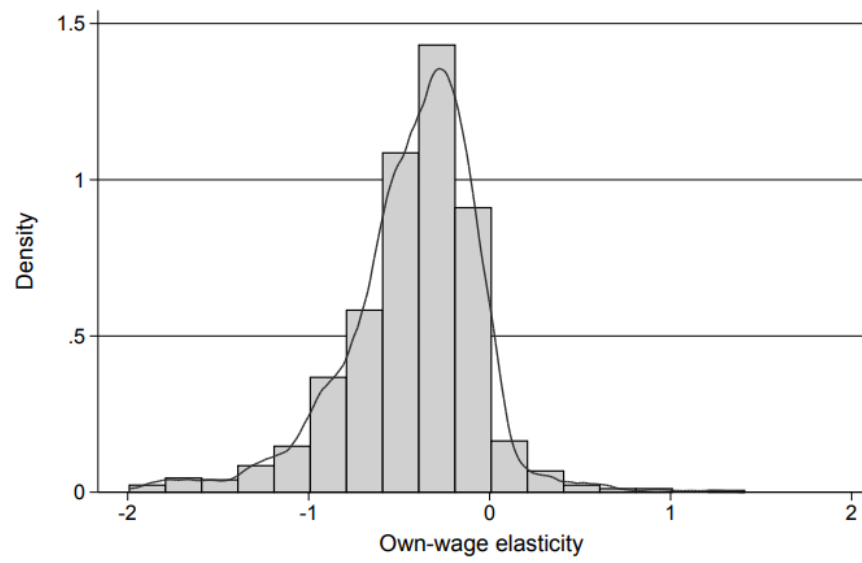
$$va^2 - 2a + 2va + 4 < 0$$

$$va^2 - 2(1 - v)a + 4 < 0$$

For  $v = 0.1$  and  $2.6 < a < 15$  this is true.

Next, for  $\gamma = 0$  the inequality  $R'(x^\gamma) > c^s$  becomes  $a > 2$  so it is satisfied for  $\gamma \approx 0$ .

Figure 1: Distribution of Labor Demand Elasticities



Taken from Lichter, Peichl and Siegloch (2015).