Labor Associations: The Blue Wall of Silence☆

David K. Levine[1], Andrea Mattozzi[2], Salvatore Modica[3]

Abstract

We develop a model showing that when labor demand is inelastic and individual behavior is easily monitored a firm's employees may prefer to protect its shirkers. By optimally reducing overall effort and increasing wages for all, a labor association rationally use its monopoly power as described in the left wing labor slogan "work less so that all may work." In addition, 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. We argue that a number of recently proposed remedies to this problem are unlikely to succeed and suggest a more promising alternative: increase competition.

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

☆First Version: March 7, 2021. We are especially grateful to Alvaro Delgado, Gilat Levy, Pietro Ortoleva, Rocco Schiavone and participants at the Madrid Political Economy Workshop, the Beijing International Workshop on Microeconomics, the Essex Theory Workshop, the UCL Theory Workshop, and the Di Tella Economics Seminar. We gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_01.

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

[1]Department of Economics and RSCAS European University Institute and Department of Economics WUSTL

[2]Department of Economics European University Institute and Università di Bologna

[3]Università di Palermo, Italy, Dipartimento di Matematica e Informatica

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 employees often protect the worst workers even when they clearly are not a majority, and what can be done about it? An obvious context evoked by the news article mentioned in the incipit is that of the police in the USA, but teachers are often blamed for similar practices.[4] Our explanation for this apparent puzzle is that when demand is inelastic effort reduction arises from the rational use of monopoly power by employees. From this directly follows that shirkers should be protected and information should not be shared with employers - what has been infamously known as the *blue wall of silence* in the case of the police.

We examine a setting in which 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 and only employees have information about their peers' effort. If labor demand is inelastic we show that it is advantageous to employees to allow shirkers to shirk, deny the firm information about employee effort, and if it is inelastic enough it is advantageous to restrain workers from working too hard. In the opposite case in which labor demand is elastic it is advantageous for employees to encourage effort and share information with the firm. Our results follow from the fact that all employees agree it would be best to maximize overall employment. When labor demand is inelastic, firms needs a certain amount of labor input "no matter what." By reducing individual effort and hence by tolerating shirking, a labor association

[4] The Mollen Commission (1994) report documents police covering up for the misbehavior of other police. Concerning teachers Moe (2011) reports *"[New York] city's Rubber Rooms – Temporary Reassignment Centers – where teachers were housed when they were considered so unsuited to teaching that they needed to be kept out of the classroom, away from the city's children. [...] They got paid a full salary. They received full benefits, as well as all the usual vacation days, and they had their summers off. Just like real teachers. Except they didn't teach."*

forces firms to hire more employees. Conversely, if demand is elastic reducing effort simply reduces the market wage.[5]

Our theory says that labor associations will protect their weakest members when demand for labor is inelastic. Given the ubiquity of protection for weakest members we must ask if it is the case that the demand for labor is typically inelastic. Indeed, we argue that the empirical evidence strongly favors inelasticity across industries and countries. In this 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 they would get if there is no labor association and the firm observes effort. In the latter case workers receive a premium because shirkers must be compensated for their effort. When the required compensation is relatively low, however, the premium will be less than the increased wage they will receive if the association's effort reduction plan is in place.

We next ask what happens if the employees can share their information with the firms and show that in the inelastic case the labor association is better off shielding the worst. The intuition is that when it is optimal to disincentivize effort, information sharing allows firms to pay incentive wages, which increases the labor association's cost of restraining workers. 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. If employees cannot prevent the leakage of information to the firm we show never-the-less that improved information will help employees but harm rather than benefit consumers. To further highlight the role of information provision by a labor association to the firm we will also consider the case where it is the firm rather than labor as-

[5]We should 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 on post probationary incentives: if shirkers learn their type during probation we would expect them to conceal it so that those laid off would most likely be types incapable of providing effort. It is also the case that wages can be the outcome of a bargaining process between firms and unions. To keep the model simple and describe our mechanism in the most transparent way we abstract from this.

sociation that receives information and provides incentives for effort. Although better information in this case will help consumers, the association will refuse to provide any additional information to the firm and, if it is able to do so, attempt to degrade information flowing to the firm.

In order to understand under what conditions labor associations can successful 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 (such as restraining effort) 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 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.[6]

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 mem-

[6]See for example Levine and Modica (2016) and Dutta, Levine and Modica (2021b).

bers. Our explanation complements existing theories which focus on particular 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.

The rest of the paper is organized as follows. We describe the model in Section 2 and derive the total cost of implementing a target level of effort and solve for the optimal mechanism in sections 3 and 4, respectively. The blue wall of silence as an equilibrium phenomenon is the focus of Section 5 and we extend our results on labor associations to individual proprietors in Section 6. Section 7 discusses the generality of our mechanism. Finally, Section 8 concludes the paper by examining possible solutions to the problem of effort provision and the blue wall of silence, focusing on the recent debate over the police in the USA. The heart of the problem in our view is one of monopoly power: the protection of shirkers and the blue wall of silence are indeed rational ways to exploit monopoly power and a "promising cure" for monopoly is competition.

2. The Model

Let x denote labor input and suppose that the value of that input to the industry $U(x)$ is strictly differentiably concave up to a satiation level X, so that the marginal value $U'(x)$ 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 ith employee is $v(i)$, which we assume to be strictly differentiably 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 γ. Letting W denote the employees' wage, when n employees are working the net utility of the ith 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 μ^s and μ^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 - subsequently we show that no mechanism can do better than these. Denote by ϕ^- a minimum quota on effort meaning that only effort levels $e \in [\phi^-, 1]$ are acceptable, and by ϕ^+ 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 π 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 ϕ and P. Average effort for an incentive compatible minimum quota is $\overline{e} = (1-\gamma)\phi^- + \gamma$ (workers provide full effort and shirkers make the minimum effort allowed) and with an incentive compatible maximum quota is $\overline{e} = \gamma\phi^+$ (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 γ and $C = 0$ since no punishment is needed in this case.

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 $n\bar{e}U'(n\bar{e}) - Wn = n\left(\bar{e}U'(n\bar{e}) - W\right)$ we have $W = \bar{e}U'(n\bar{e})$. On the supply side it must be $W - C \leq v(n)$ 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 $\hat{n}(\gamma, 0) > 0$ is uniquely defined by $\gamma U'(\gamma\hat{n}(\gamma, 0)) = v(\hat{n}(\gamma, 0))$. It will be convenient to define $x^\gamma = \gamma\hat{n}(\gamma, 0)$, which is the market clearing output when workers work and shirkers shirk. We refer to γ as the *natural level of effort* and x^γ as the *natural level of labor input.* An important fact established in the next Lemma is that x^γ is increasing in γ.

Lemma 1. *The natural level of labor input and corresponding consumer welfare are strictly increasing in γ.*

Proof. Since consumer welfare is $U(x^\gamma)$ it suffices to show $\partial x^\gamma / \partial\gamma > 0$. From the implicit function theorem

$$\frac{\partial \gamma\hat{n}}{\partial\gamma} = -\frac{U'(\hat{n}\gamma) + \gamma\hat{n}U''(\hat{n}\gamma)}{\gamma^2 U''(\hat{n}\gamma) - v'(\hat{n})}\gamma + \hat{n}$$

$$= -\frac{U'(x^\gamma)}{\gamma} - \frac{\gamma x^\gamma U''(x^\gamma)}{\gamma^2 U''(x^\gamma) - v'(\hat{n})} + \hat{n}$$

$$= -\frac{U'(x^\gamma)}{\gamma} - \frac{(x^\gamma)^2 U''(x^\gamma)}{(x^\gamma)^2 U''(x^\gamma) - v'(\hat{n})}\hat{n} + \hat{n} > 0.$$

□

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 γ 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 $\overline{e} < \gamma$, 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 $\overline{e}$ is given by*

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

From this we see that the marginal implementation costs c^w and c^s are increasing in μ^w and μ^s respectively, and both are increasing in the monitoring difficulty θ. On the other hand c^w decreases in γ while c^s increases in γ.

Proof. Incentive compatibility for a minimum quota is that shirkers must prefer providing ϕ^- 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 $\overline{e} = \gamma$ 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 $\overline{e} > \gamma$, a minimum quota must be established so that $\overline{e} = (1-\gamma)\phi^- + \gamma$, with corresponding cost $C = \mu^s(1-\gamma+\theta)\phi^- = \mu^s(1-\gamma+\theta)(\overline{e}-\gamma)/(1-\gamma)$. If $\overline{e} < \gamma$, a maximum quota must be established so that $\overline{e} = \gamma\phi^+$, with corresponding cost $C = \mu^w(\gamma+\theta)(1-\phi^+) = \mu^w(\gamma+\theta)(1-\overline{e}/\gamma) = \mu^w(\gamma+\theta)(\gamma-\overline{e})/\gamma$. □

4. The Optimal Mechanism

Slightly abusing notation we let $\hat{n}(\overline{e}) \equiv \hat{n}(\overline{e}, C(\overline{e}))$. An optimal target level of average effort is a choice $\overline{e} = \hat{e}$ which maximizes $\hat{n}(\overline{e})$. We let $\hat{x} = \hat{n}(\hat{e})\hat{e}$ so that $\hat{x}$ denotes the *optimal* level of labor input as opposed to x^γ, which is the *natural* level of labor input. Finally, recall that $R(x) = xU'(x)$ is the (concave) revenue function. We are now ready to state our first main result. Recall that since consumer welfare is $U(\hat{x})$, it has the same comparative static as $\hat{x}$.

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

- (low) If $R'(x^{\gamma}) < -c^{w}$ then a maximum quota is optimal, $\hat{e} < \gamma$, $\hat{x} < x^{\gamma}$, and $R'(\hat{x}) = -c^{w}$. Furthermore, optimal labor input and consumer welfare are increasing in c^{w}, while employee utility is strictly decreasing in μ^{w} and θ;

- (natural zone) 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 a minimum quota is optimal ,$\hat{e} > \gamma$, $\hat{x} > x^{\gamma}$and $R'(\hat{x}) \geq c^{s}$ with equality if $\hat{e} < 1$. Furthermore, if $R'(\hat{x}) = c^{s}$ optimal labor input and consumer welfare are decreasing in c^{s} and c^{w}, while employee utility is strictly decreasing in μ^{s} and θ.

Before proving this central result of the paper three observations are worth mentioning. First, since c^{w} and c^{s} increase in θ, for large enough θ - that is an inefficient monitoring technology - we are in the natural zone and therefore setting a quota is definitely not optimal. This is fairly intuitive since if monitoring is very difficult then implementing any level of effort different from the natural level γ is extremely costly. Second, notice that if $R'(\hat{x}) \leq 0$ then we must either be in the natural zone or in the low domain. Since marginal revenue is negative if and only if demand is inelastic, we can rephrase this by saying that 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.

Finally, notice that increasing γ has two effects on labor input. It increases x^{γ} by Lemma 1 and it decreases c^{w} and increases c^{s} shortening the width of the natural zone. One implication of this is that increasing γ moves x^{γ} towards the low zone, that is, with many workers we are more likely to be in the inelastic case. This means that the idea that workers outnumber shirkers is consistent with the inelastic case. Furthermore, the effect of γ on labor input and consumer welfare is not monotone. In the high zone increasing c^{s} must increase $\hat{x}$ and in the natural zone $\hat{x} = x^{\gamma}$ also increases. In the low zone increasing γ decreases c^{w} which decreases $\hat{x}$. In other words, labor input and consumer welfare increases with γ until the lower end of the natural zone is reached then it begins to decline. While it is natural to think that more productive workers are good for consumers, we see that in the inelastic case this is not true: increasing γ reduces the marginal cost of reducing effort, and the reduction of effort leads to a decrease in labor input.

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

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

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

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

where we used that $C''(\overline{e}) = 0$ from Theorem 1. This implies that $R'(\hat{n}\overline{e}) - C'(\overline{e}) = 0$ is always a local maximum and not a local minimum. From Theorem 1, we know that in the lower domain $R'(\hat{n}\overline{e}) - C'(\overline{e}) = R'(\hat{n}\overline{e}) + c^w$ while in the higher domain $R'(\hat{n}\overline{e}) - C'(\overline{e}) = R'(\hat{n}\overline{e}) - c^s$. Hence in the lower domain if $R'(x^\gamma) + c^w \geq 0$ there can be no local maximum with $\overline{e} < \gamma$, while in the higher domain if $R'(x^\gamma) - c^s \leq 0$ there can be no local maximum with $\overline{e} > \gamma$. Hence if both these conditions hold we are in the natural zone. 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 $R'(\hat{n}\hat{e}) = -c^w < 0$.

The comparative static results about $\hat{x}$ follow directly from the first order conditions and the fact that marginal revenue is assumed to be decreasing. The comparative statics about employee utility follows from the fact that the increases strictly lower the objective function. □

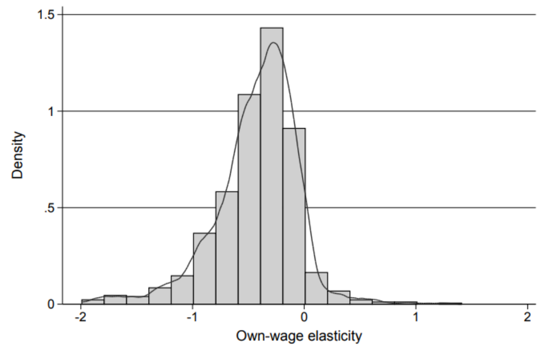
How Elastic is the Demand for Labor?

Our theory says that labor associations will protect their weakest members, but they will do so exactly when demand for labor is inelastic. Given the ubiquity of protection for weakest members we must ask if is indeed the case that the demand for labor is typically inelastic. Lichter, Peichl and Siegloch (2015) do a meta-study of labor demand. We reproduce below their figure showing estimates from different studies of the elasticity of labor demand.[7] As can be seen the vast bulk of estimates lie above -1, so correspond to inelastic demand.

---

[7]Notice that the studies that underlie this data refer to elasticity at the equilibrium not at the natural level, while the theory does the opposite. However, Theorem 2 shows that if demand is elastic at x^γ it is at $\hat{x}$ and conversely.

Figure 1: Distribution of Labor Demand Elasticities



We should observe that labor demand studies measure the elasticity of hours (n) with respect to wages not the elasticity of labor input with respect to wages (x) which is what our theory refers to. In our theory effort is endogenous so these two elasticities are not necessarily the same. In particular, there are two cases to consider. If we are in the natural zone or we observe either the short run or shocks are temporary, effort is fixed and the two elasticities are indeed the same.[8] If this is not the case labor input $\hat{x}$ does not respond to shocks to labor supply v and these shocks will not change wages as these are on the demand curve. They will however change employment n: this is met by the labor association by adjusting effort to keep labor input fixed. That is, shocks will move employment but not wages. In this case measured elasticity is infinite.

As a practical matter we may suppose that labor demand studies observe something in between the very short run and the very long run so that the measured workforce elasticity will be more elastic than labor input elasticity, although not actually infinite. Since measured workforce elasticity is generally greater than -1 we may conclude the same is true for labor input elasticity and correspondingly $R'(x^\gamma) < 0$.

[8] In the natural zone effort is fixed at γ while in the the short run or if shocks are temporary it is reasonable to suppose that the labor association is unable to adjust quotas so that again effort is fixed.

Utility of Workers

To what extent are workers content with their colleagues shirking? Specifically we analyze the case in which $-c^w \leq R'(x^\gamma) < 0$ so that the labor association is passive and workers work and shirkers shirk. 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 consider the following conceptual experiment. First, suppose that firms perfectly observes effort but are prohibited by a union contract from paying incentive wages. Second, suppose that after employment and after employees learn their type a vote is taken over whether to keep the labor association or to disband the labor association and allow the firm to pay incentive wages. Define n^* as the competitive equilibrium with full effort: $u'(n^*) = v(n^*)$.

Theorem 3. *Suppose that $-c^w \leq R'(x^\gamma) < 0$ and $\mu^s < U'(n^*)$. Then there exists an $m > 0$ such that for $0 < (1-\gamma)\mu^s < m$ workers are strictly better off with a labor association.*

Notice in particular that the condition $(1-\gamma)\mu^s$ is small will be satisfied if γ is large - that is, if there are many workers they will be strictly better off with the labor association.

Proof. Suppose there is no labor association and the firm can observe effort. For $\mu^s < U'(n^*)$, as in any equilibrium $n \leq n^*$ we have $\mu^s < U'(n)$ and it is efficient for shirkers to provide full effort. Hence the equilibrium wage (with full effort) is $W^{NA} = U'(n^{NA}) = v(n^{NA}) + \mu^s(1-\gamma)$ where n^{NA} is equilibrium employment. That is, workers get a premium because shirkers must be compensated for their effort. With a labor association, since $\overline{e} = \gamma$ and $C = 0$, the wage is given by $W^A = \gamma U'(\gamma n^A) = v(n^A)$. Let $W^* = U'(n^*)$. Because $R' < 0$ it must be that $n^A > n^*$ and $W^A > W^*$. As n^*, W^* corresponds to n^{NA}, W^{NA} with $\mu^s = 0$, by continuity for small enough μ^s we have $n^A > n^{NA}, W^A > W^{NA}$.

To assess the situation for a worker, observe that while without a labor association employment is reduced in the industry there are not layoffs: workers who leave the industry do so voluntarily. Since in the natural zone workers are able to provide full effort and there is no punishment, the relevant consideration for a worker is their wage if they stay in the industry - this is strictly higher if there is a labor association so they are strictly better off. $\square$

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.

An alternative way to analyze the issue of worker attitudes towards shirkers is this: if shirkers by reducing overall effort increase utility for workers then we imagine that workers are grateful to shirkers. Our next result shows that this is in fact the case.

Theorem 4. *Suppose that $-c^w \leq R'(x^\gamma) < 0$. Then worker utility is decreasing in γ.*

This theorem resolves a phenomenon that has long puzzled us. We have observed, for example, in the Italian Post Office, at Departments of Motor Vehicles, and in the private sector at rental car agencies, long queues and a number of windows for servicing customers. Behind most of these windows are employees shuffling papers or otherwise shirking.[9] Behind one window is an employee working like a demon trying to get the customers what they want. The question is why the worker puts up with the shirkers. As it appears that demand is generally inelastic, the answer we get from our theory is that the worker - who likes to work - receives a higher utility due to the presence of the shirkers.

Proof. Define $\hat{n} = \hat{n}(\gamma, 0)$.From the market clearing condition $\gamma U'(\gamma\hat{n}) = v(\hat{n})$

[9]An interesting example is provided in the Walt Disney movie Zootropolis where all of DMV employees are sloths executing tasks extremely slowly, much to the frustration of customers.

and the implicit function theorem we get

$$\frac{\partial \hat{n}}{\partial \gamma} = -\frac{U'(\hat{n}\gamma) + \gamma \hat{n} U''(\hat{n}\gamma)}{\gamma^2 U''(\hat{n}\gamma) - v'(\hat{n})} = -\frac{R'(x^\gamma)}{\gamma^2 U''(x^\gamma) - v'(\hat{n})},$$

which is negative when $R'(x^\gamma) < 0$. Since in the natural zone workers get to make full effort and there is no punishment, their utility is their wage, and this decreases as higher γ lowers employment. $\square$

5. The Blue Wall of Silence

We now turn to the role of the firm, and specifically whether the labor association is willing to provide information to the firm about employee behavior. In particular we now suppose 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 ϕ, 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.

Our basic assumption is that the firm cannot impose unlimited wage penalties, but faces a constraint in how much it can penalize bad signals. The expected wage paid by the firm is $\pi W(0) + (1-\pi)W(1) = W$. We make the relatively general assumption that $-\underline{\Delta}(W) \leq W(1) - W(0) \leq \overline{\Delta}(W)$ where $\overline{\Delta}(W), \underline{\Delta}(W) > 0$ and are continuous in W. For example, if the firm is limited to non-negative wages so that $W(z^i) \geq 0$ then $\overline{\Delta}(W) = W/(1-\pi)$ and $\underline{\Delta}(W) = W/\pi$.

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 π' by exceeding the quota by any positive ϵ. 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 ϵ then, if they exactly meet the quota, they can choose between π and π', that is they can choose if they wish the probability of generating a bad signal. Given this we are ready to state our second main result.

Theorem 5. *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 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 ϕ, P such that there exists a firm optimal choice $\pi W(0) + (1-\pi)W(1) = W$ that is incentive compatible for both types. The premium for a good signal is $\Delta =$

$W(1) - W(0)$. In the case of a minimum quota the binding constraint is that shirkers must prefer providing ϕ^- to not providing effort. This is $\mu^s\phi^- + \pi P - \pi W(0) - (1-\pi)W(1) \leq \pi' P - \pi' W(0) - (1-\pi')W(1)$, or

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

For the maximum quota analogously we have for workers

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

Observe that a deviation from Δ to Δ' 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 $P = \max\{0, \mu^s\hat{\phi}^-/(\pi'-\pi) - \overline{\Delta}(\hat{W})$, where - recall that the punishment value in the original model is $\mu^s\hat{\phi}^-/(\pi'-\pi)$. If the firm chooses $W = \hat{W}$ and $\Delta = \overline{\Delta}(\hat{W})$ this is optimal for the firm since Δ can be increased only by raising W which raises firm costs without changing worker behavior;[10] and decreasing Δ will induce the shirkers to violate the constraint decreasing output. Let $x = n\left(\gamma + (1-\gamma)\hat{\phi}^-\right)$. The reduction in output costs the firm $U'(x)(1-\gamma)\hat{\phi}^-$ but enables it to reduce wages by $\mu^s(1-\gamma)\hat{\phi}^-$. We know that

$$U'(x) \geq R'(x) > c^s = \mu^s \frac{1-\gamma+\theta}{1-\gamma} \geq \mu^s$$

so this is unprofitable. 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

[10] Note that the firm cannot decrease W since then it will lose all its workers.

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 Δ 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. $\square$

Involuntary Disclosure in the Inelastic Case

We now examine the situation in which the labor association cannot prevent the firm from observing the signal. We are particularly interested in how the comparative statics change in this case. For simplicity we will assume that the maximum wage differentials $\overline{\Delta}(W), \underline{\Delta}(W)$ are a constants $\underline{\Delta}, \overline{\Delta}$ independent of W.

Theorem 6. *Suppose that $R'(x^\gamma) \leq 0$ and define $\tilde{v}(n) = v(n) + \pi\underline{\Delta}$. Then the equilibrium in which the labor association cannot prevent the firm from observing* *the signal is the same as that in which the firm does not observe the signal and opportunity cost is given by $\tilde{v}(n)$. Labor association utility is decreasing in $\pi\underline{\Delta}$ and consumer utility increasing.*

This implies in particular that in the inelastic case improved information (lower π) is better for the labor association but worse for consumers. In the case of the police improved monitoring technology such as body cams may in fact reduce consumer welfare.

Proof. Note that since the labor association controls the quota they need not accept any output higher than the natural level x^γ since they can attain this at zero cost by setting a maximum quota of zero. The incentive constraint is $\mu^w(1-\phi^+) \leq (\pi' - \pi)(P + \Delta)$ and must hold for all $\Delta \geq -\underline{\Delta}$ so is equivalent to $\mu^w(1-\phi^+) \leq (\pi' - \pi)(P - \underline{\Delta})$. Minimizing with respect to P gives

$$P = \frac{\mu^w(1-\phi^+)}{\pi' - \pi} + \underline{\Delta}$$

resulting in a cost of $C(\overline{e}) = -\left(\mu^w(\gamma+\theta)/\gamma\right)(\overline{e} - \gamma) + \pi\underline{\Delta}$. In the equilibrium condition this is equivalent to shifting the opportunity cost of labor up by $\pi\underline{\Delta}$.

In the natural zone utility is unchanged since the signal is being used. When the first order condition holds $R'(\hat{n}\hat{e}) + c^w = 0$ and the equilibrium condition is $U'(\hat{n}\hat{e}) = v(\hat{n}) + \pi\underline{\Delta}$. Since $\hat{x}$ does not depend on $\pi\underline{\Delta}$ we see that consumer welfare does not change, and from the implicit function theorem follows that

$$\frac{\partial \hat{n}}{\partial \pi\underline{\Delta}} = -\frac{1}{v'(\hat{n})} < 0$$

meaning that the labor association is worse off.

However: increasing $\pi\underline{\Delta}$ will cause the solution to $\gamma U'(n\gamma) = v(n) + \pi\underline{\Delta}$ to decline which can cause a jump from the low solution to the natural zone: this would raise consumer utility (and lower labor association utility). □

Strong Firms

To further highlight the role of information provision by a labor association to the firm we will consider the case where the firm rather than the labor association sets the quota (which we may assume is a minimum quota) and receives the signal. For simplicity we continue to examine the case in which the maximum wage differentials $\overline{\Delta}(W), \underline{\Delta}(W)$ are a constants $\underline{\Delta}, \overline{\Delta}$ independent of W. Our interest is in how *signal quality*, measured by $\sigma = \pi' - \pi$ impacts on consumers and the association. Even in the absence of provision of information by the association we may have $\sigma > 0$: for example in the case of the police, civilians with cell phone cameras and body cams may provide useful information about police behavior. While the association does not control the quota, cannot punish, and perhaps does not even see the signal, it can improve the quality of the signal by providing information to the firm. We will establish that greater σ increases labor input and hence consumer utility, but that in the inelastic case where $R'(x^\gamma) < 0$ it reduces employment and hence the association's utility $v(n)$. In this case the association will refuse to provide information to the firm and indeed, if it is able to do so - for example, in the case of the police, by harassing civilian photographers and sabotaging body cameras - the association will attempt to degrade that information. In conclusion, a better firm signal improves consumer welfare but does not break the blue wall of silence.

For concreteness we state the problem of a representative firm and the equilibrium conditions, and we focus on minimum quotas. To ease notation we will suppress the superscript and write simply ϕ to denote the quota ϕ^-. The firm pays $W(0)$ for a bad signal and $W(1) = W(0) + \Delta$ for a good signal and sets an incentive compatible minimum quota $1 \geq \phi \geq 0$. The wage differential constraint is $-\underline{\Delta} \leq \Delta \leq \overline{\Delta}$. Since only the minimum quota is being used the lower bound does not matter, so we may work with the constraint $\Delta \leq \overline{\Delta}$. Workers work and the quota must be incentive compatible for shirkers $\mu^s\phi \leq (\pi' - \pi)\Delta$. The wage bill per worker for the firm is $W = \pi W(0) + (1 - \pi)W(1) = W(0) + (1 - \pi)\Delta$. The utility provided to a worker is $\nu = W - (1-\gamma)\mu^s\phi$. In the market the firm takes as given the output price, which we denote by Q, and the worker utility which we denote by V so that it maximizes per worker profits $Q((1-\gamma)\phi+\gamma)-W$ subject to $\Delta \leq \overline{\Delta}$, the incentive constraint and $\nu \geq V$. In equilibrium $(1-\gamma)\phi+\gamma = \hat{e}$, $\nu = V = v(n)$, $Q = U'(n\hat{e})$ and there is zero profit per worker.

Define *high effort* by $\hat{\phi} = \min\{1, \sigma\overline{\Delta}/\mu^s\}$, $\overline{x}$ as the unique solution of $U'(\overline{x}) = \mu^s$, $\overline{n}$ as the unique solution to $v(n) = \gamma\mu^s$ and $\overline{\phi} = (\gamma/(1-\gamma))\overline{x}/\overline{n}$. Our key result characterizes the equilibrium.

Theorem 7. *There is a unique equilibrium given as follows:*

(low effort) If $\overline{\phi} < 0$ then $\phi = 0$, employment n^ℓ is the unique solution of $\gamma U'(n\gamma) - v(n) = 0$ with labor input $x^\ell = \gamma n^\ell$.

(intermediate effort) If $0 \le \overline{\phi} \le \hat{\phi}$ then $\phi = \overline{\phi}$, employment is $\overline{n}$ with labor input $x^m = \big((1-\gamma)\overline{\phi} + \gamma\big)\,\overline{n}$.

(high effort) If $\overline{\phi} > \hat{\phi}$ then $\phi = \hat{\phi}$, employment n^h is the unique solution of

$$U'\left(n((1-\gamma)\hat{\phi}+\gamma)\right)((1-\gamma)\hat{\phi}+\gamma) - v(n) - \mu^s(1-\gamma)\hat{\phi} = 0$$

with labor input $x^h = \left((1-\gamma)\hat{\phi}+\gamma\right) n^h$.

For fixed γ, μ^s employment is strictly increasing in ϕ for $R'(x) - \mu^s > 0$ and strictly decreasing for $R'(x) - \mu^s < 0$, while labor input and consumer welfare are strictly increasing in both cases.

To interpret this result, observe that increasing σ increases $\hat{\phi} = \min\{1, \sigma\overline{\Delta}/\mu^s\}$ when it is less than 1. In the low and natural case this has no effect on the equilibrium, but in the high effort case with $\sigma < \mu^s/\overline{\Delta}$ it increases ϕ hence increases labor input. Moreover, $x^h \ge x^\gamma$ since x^γ corresponds to $\phi = 0$ and since in the inelastic case $R'(x^\gamma) < 0$ it must be that $R'(x^h) - \mu^s < 0$. This means that employment and employees' utility are strictly decreasing in the signal quality: the labor association will not provide additional information to the firm, and if it is able to do so will degrade the information received by the firm.

Proof. We may write the per worker profit as $Q((1-\gamma)\phi+\gamma) - W(0) - (1-\pi)\Delta$ and this must be maximized subject to $\Delta \le \overline{\Delta}$, $W(0) + (1-\pi)\Delta - (1-\gamma)\mu^s\phi \ge V$, $(\pi' - \pi)\Delta \ge \mu^s\phi$, and $0 \le \phi \le 1$.

Suppose we have a solution to this problem $W(0), \Delta, \phi$ and consider an alternative profile $\tilde{W}(0) = W(0) - (1-\pi)(\overline{\Delta} - \Delta)$, $\tilde{\Delta} = \overline{\Delta}$ and $\tilde{\phi} = \phi$: this is feasible and yields the same profit so it also optimal. Hence we may assume $\Delta = \overline{\Delta}$. Hence per worker profit is $Q((1-\gamma)\phi + \gamma) - W(0) - (1-\pi)\overline{\Delta}$

and this must be maximized subject to $W(0) + (1-\pi)\overline{\Delta} - (1-\gamma)\mu^s\phi \geq V$, $(\pi' - \pi)\overline{\Delta} \geq \mu^s\phi$, and $0 \leq \phi \leq 1$.

We immediately see that the constraint $W(0) + (1-\pi)\overline{\Delta} - (1-\gamma)\mu^s\phi \geq V$ must bind so $W(0) = V + (1-\gamma)\mu^s\phi - (1-\pi)\overline{\Delta}$ and profits are

$$Q((1-\gamma)\phi + \gamma) - V - (1-\gamma)\mu^s\phi.$$

The derivative with respect to ϕ is $Q(1-\gamma) - (1-\gamma)\mu^s$. If $Q < \mu^s$ then $\phi = 0$. If $Q = \mu^s$ then ϕ is limited by the second constraint $(\pi' - \pi)\overline{\Delta} \geq \mu^s\phi$. If $Q > \mu^s$ then the constraint on ϕ must bind. Hence we have that:

If $Q < \mu^s$ then $\phi = 0$ and in equilibrium $Q = U'(\gamma)$ and profits are $\gamma U'(\gamma) - v(n)$ which is positive for $n < n^\ell$ and negative for $n > n^\ell$.

If $Q = \mu^s$ then profits are $\mu^s((1-\gamma)\phi + \gamma) - v(n) - (1-\gamma)\mu^s\phi = \mu^s\gamma - v(n)$ which is is positive for $n < \overline{n}$ and negative for $n > \overline{n}$.

If $Q > \mu^s$ then the constraint on ϕ must bind so that $\phi = \hat{\phi}$ and $W(0) = V + (1-\gamma)\mu^s\hat{\phi} - (1-\pi)\overline{\Delta}$. Take $\hat{e} = \gamma + (1-\gamma)\hat{\phi}$. Per worker profit is

$$Q\hat{e} - \left[V + (1-\gamma)\mu^s(\hat{e} - \gamma) - (1-\pi)\overline{\Delta}\right] - (1-\pi)\overline{\Delta}$$

$$= Q\hat{e} - V - \mu^s(\hat{e} - \gamma).$$

The equilibrium condition is then

$$U'(n\hat{e})\hat{e} - v(n) - \mu^s(\hat{e} - \gamma) = 0.$$

In all cases profits are zero

$$U'(n((1-\gamma)\phi + \gamma))((1-\gamma)\phi + \gamma) - v(n) - (1-\gamma)\mu^s\phi = 0.$$

From the implicit function theorem

$$\frac{dn}{d\phi} = -(1-\gamma)\frac{((1-\gamma)\phi + \gamma)nU''(x) + U'(x) - \mu^s}{((1-\gamma)\phi + \gamma)^2 U''(x) - v'(n)}$$

where $x = ((1-\gamma)\phi + \gamma)n$, which further simplify into

$$= -(1-\gamma)\frac{R'(x) - \mu^s}{((1-\gamma)\phi + \gamma)^2 U''(x) - v'(n)}$$

Moreover,

$$\frac{dx}{d\phi} = (1-\gamma)n + ((1-\gamma)\phi + \gamma)\frac{dn}{d\phi}$$

$$= (1-\gamma)\left(-\frac{xU''(x) + U'(x) - \mu^s}{((1-\gamma)\phi + \gamma)^2 U''(x) - v'(n)}((1-\gamma)\phi + \gamma) + n\right)$$

$$= (1-\gamma)\left(\frac{x^2 U''(x)n + (U'(x) - \mu^s)nx}{-x^2 U''(x) + n^2 v'(n)} + n\right)$$

Since

$$\frac{x^2 U''(x)n}{-x^2 U''(x) + n^2 v'(n)} > -n$$

we have $dx/d\phi > 0$ for $U'(x) \geq \mu^s$, and x, ϕ constant for $U'(x) < \mu^s$.

We conclude that $x^h > x^m > x^l$ hence $U'(x^h) < U'(x^m) < U'(x^l)$. If $\mu^s < U'(x^h)$ then it is the high case and $dx/d\phi > 0$ implies $d\phi/dx > 0$ so $\overline{\phi} > \hat{\phi}$. If $U'(x^h) \leq \mu^s \leq U'(x^l)$ then by definition of x^m we have $U'(x^m) = \mu^s$ putting us in the intermediate case and $d\phi/dx \geq 0$ implies that $0 \leq \overline{\phi} \leq \hat{\phi}$. Finally if $\mu^s > U'(x^l)$ then we are in the low case and clearly $\phi = 0$. $\square$

6. 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 $\overline{e}$ total output is $x = N\overline{e}$. We continue to assume the value of output $U(x)$ is strictly differentiably concave up to a satiation level $X > N$ and that the revenue function $R(x) = xU'(x)$

is concave. Now, however, proprietors face a constant marginal cost ξ of other inputs used in producing output and are paid individually for the output they produce, so that the profit of a proprietor is $U'(x)e - \xi e$. We assume that $U'(N) > \xi$ so that the market clearing effort level absent any incentives is N with corresponding market price $U'(N)$. In this context $U'(x)/xU''(x)$ 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 ϕ^- or a maximum quota ϕ^+. 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*

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

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

Theorem 8. *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 $U'(N\overline{e})e - \xi e$. Since $\overline{e} \leq 1$, $U'(N) > \xi$ 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 $U'(N\phi^+)\phi^+ - c\phi^+ - \pi P \geq U'(N\phi^+) - \xi - \pi' P$ or

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

Hence profits of a member of the association is given by

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

Differentiate with respect to ϕ^+ to get

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

This has the same sign as

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

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

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

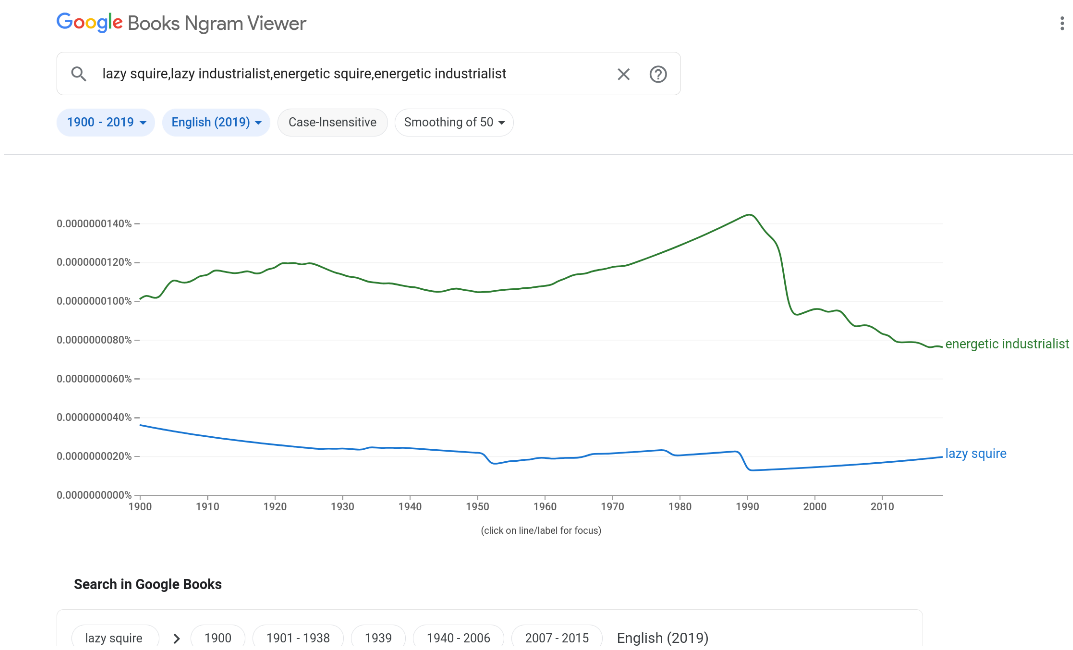
□

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 8 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 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 8 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.

7. General Mechanisms

Our analysis has been of a special class of mechanisms: a quota with a bad punishment for a bad signal. Could the labor association do better with a more general mechanism? Roughly speaking the answer is no, but to make this precise we need to consider carefully what a general mechanism would look like.

We should indicate first that the "quota plus signal" is a special case of the type of flexible information system studied by Yang (2015). That is, we may define a flexible class of effort dependent *information systems* each defined by a threshold φ and a direction φ^+, φ^-. If effort lies at or below φ^+ the signal 0 is emitted with probability π and if it lies above the signal 0 is emitted with probability π'. Similarly the information systems φ^- emit 0 for effort at or above φ^- with probability π and below with probability π'. Here the information systems exhibit high sensitivity near the threshold. In Yang (2020) information systems for designing a bond should be sensitive near the default boundary: here they should be sensitive near the target effort level. Indeed, due to the discontinuity at the threshold the only incentive compatible effort targets are either at the discontinuity - equivalent to our quota model - or at 0 or 1 - which is also equivalent to our quota model.[11]

Next, in addition to choosing an information system, the labor association can ask members to reveal their types. It can then issue punishments based on the combination of type statements and signals. This would be the "general mechanism." An important issue is whether in addition to type contingent punishments the association can choose a type contingent information system. For example, it may be that the information system has to be chosen before types are realized. Implicitly we have assumed that this is the case; we indicate below what happens if the association can choose type-contingent information systems.

In the case of non-type contingent information systems, what extra leverage does the labor association gain from punishments that are type and signal contingent rather than merely punishing based on a bad signal? In this discussion, bear in mind that the relevant consideration is how the incentive constraints impact on the cost of achieving an effort target $\bar{e}$. The answer depends on whether or not $\pi < 1 - \pi'$. The reason is that (allowing employees to choose which signal probability to use when they are on the effort boundary $e = \varphi$) if $\pi > 1 - \pi'$ off path punishment costs could be reduced by reversing the role of the two information systems so that the on-path punishment probability would be $1 - \pi'$ rather than π. On the one hand this is really notational, since we can

[11]More general information systems including continuous ones are studied in Dutta, Levine and Modica (2021a) who show that if the sensitivity is large enough equilibrium choice of effort resembles that in the discontinuous case.

just redefine the probabilities accordingly. Moreover, an issue not yet studied in the flexible information system literature is that of evasion: the signal that receives punishment creates incentives for the employee to obscure the signal. Hence it might be that $1-\pi'$ is relatively large because employees try to conceal their bad signals.

Assuming either that $\pi < 1 - \pi'$ or that reversal is impossible, basing punishments on types will simply cause employees to lie about their type to receive the lesser punishment. Similarly it makes no sense to punish on both signals since this reduces incentive compatibility while increasing cost. Hence we conclude that the mechanism studied here is indeed the best in the class of general mechanisms.

If it is possible to base the information system on type revelation then the model changes to one with a type-contingent quota. Each quota can have its own punishment which we may denote by P^τ where τ is the type. Let P denote the cost minimizing punishment in the original model with a type independent information system. We have

Theorem 9. *Cost minimization implies $P^s = P^w = P$, and in particular $C(\overline{e})$ does not depend on whether or not type dependent information systems are available.*

Proof. In the inelastic case take $\tau = w$ and in the elastic case take $\tau = s$. For the given maximum or minimum quota we must still minimize cost and have incentive compatibility, meaning that $P^\tau = P$. Certainly $P^{-\tau} = P$ is feasible and $P^{-\tau} > P$ raises costs, so we only need to show that $P^{-\tau} < P^t$ is not incentive compatible. To see why, notice that type τ must be indifferent to their favorite effort level using their own information system and punishment, so receives a utility of $V^\tau \geq -\pi' P^\tau$. If they lie about their type and choose their favorite effort level they would instead get $-\pi' P^{-\tau}$. Hence truth-telling requires $P^{-\tau} \geq V^\tau/\pi' \geq P^\tau$. □

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? In the context of the police three strategies have been suggested. The first is to abolish police unions. The second is the increased provision of information - laws that prevent police from interfering with civilians recording encounters and mandating the use of body cams. The third is to "defund the police." Based on our model each of these strategies is problematic; the more traditional solution to monopoly - competition - seems more promising.

Clearly police unions are not a problem per-se. Rather it is the social network of police officers enabling monitoring and peer punishment that leads to effort reduction and the blue wall of silence. Whether there is a formal structure - a union - or not, the police can engage in informal discouragement of effort, and indeed Ostrom (1990) clearly documents how formal institutions are not needed for collective action.

Consider, second, the increased provision of information to firms. This is a double-edged sword. It enables firms to provide better incentives to work - but it also reduces the cost to labor associations of discouraging work. When the labor association controls the quotas we showed that better information makes consumers strictly worse off.[12] In the case of strong firms it does improve consumer welfare but it does not give the labor association any incentive to provide additional information. Hence, it cannot crack the blue wall of silence and we may expect the labor association to fight back by trying to reduce the flow of information.

The policy of "defund the police" is not always clearly described. One thing it may mean is replacing the police with a different type of police with different or additional training, perhaps mental health workers. Since the incentives of whatever association provides "policing" services are the same, it is hard to see how this helps. Alternatively "defund the police" may 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

[12]This may explain why some police unions have favored body cams: see, for example, https://eu.statesmanjournal.com/story/news/crime/2020/08/03/police-reform-salem-oregon-brutality-body-camera-budget/5453765002/.

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.[13] 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.

The heart of the problem in our view is one of monopoly power: the protection of shirkers and the blue wall of silence are rational ways to exploit monopoly power. In general the "cure" for monopoly is competition.[14] In the model this corresponds to increasing the elasticity of demand. Not only does this increase labor input and consumer welfare but it can potentially break the blue wall of silence entirely - as we showed, with enough demand elasticity the labor association prefers to provide information to the firm.

How can increased competition come about? Consider breaking the "defund the police" scheme into specific police services: investigation, patrolling, response to domestic incidents and so forth. Narrower product categories generally have greater elasticity of demand. Hence, rather than a single "one-size fits all" police force, each of these services could be contracted to a different provider. For example, in the USA, the FBI could be hired to investigate, a private security service to patrol, and a mental health provider to respond to domestic incidents. If the social networks of these different providers are different, competition is induced between labor associations. Hence traditional police forces might bid against mental health firms for the contract to respond to domestic incidents - and police social networks are rather different than those of mental health firms. This increased competition also increases the elasticity of demand. While this solution has somewhat the flavor of "defund the police" it is better described as "make them compete."

[13] https://cops.usdoj.gov/pdf/taskforce/taskforce_finalreport.pdf

[14] There is evidence that increased competition improves labor productivity: see for example Galdon-Sanchez and Schmitz Jr (2002).

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 (2021a): "Interventions with Sticky Social Norms: A Critique," forthcoming *Journal of the European Economic Association.*

Dutta, R., D. K. Levine and S. Modica (2021b): “The Whip and the Bible: Punishment Versus Internalization,” forthcoming *Journal of Public Economic Theory.*

Fielding, Henry (1742): *Joseph Andrews.*

Galdon-Sanchez, J. E., and J. A. Schmitz Jr (2002): “Competitive pressure and labor productivity: World iron-ore markets in the 1980’s,” *American Economic Review* 92: 1222-1235.

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, Lutchmie 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.

Commision to Investigate Allegations of Police Corruption and the Anti-corruption Procedures of the Police Department Report. The City of New York, New York.

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.

Yang, M. (2015): "Coordination with flexible information acquisition," *Journal of Economic Theory* 158: 721-738.

Yang, M. (2020): "Optimality of Debt under Flexible Information Acquisition," *Review of Economic Studies* 87: 487-536

Zieger, R. H. (1994): *American workers, American unions*, Johns Hopkins Univ. Press.