MIT Graduate Labor Economics 14.662 Spring 2015
Lecture Note 6: Institutions, Norms, Collective Bargaining and Worker Productivity

David H. Autor
MIT and NBER

March 16, 2015

1 Introduction

In contrast to the neoclassical conception of the labor market as a frictionless spot market, there are a very large number of institutions that shape the supply and demand for labor, the nature of contracting, the provision of pay and benefits, etc. I group labor market institutions into roughly four categories, though there is no standard practice here, and many categories overlap:

1.1 Contracting, legal and collective bargaining environment

- What is the default contracting regime, such as Employment at Will or Master and Servant.
- Do workers have a right to collectively bargain? Do firms?
- Are there 'right to work' laws that undercut collective bargaining?
- Is there a legal authority that oversees union activity and elections?
- Is there an arbitration regime?
- How large or small, centralized or decentralized are labor unions?
- What can and cannot be bargained over (rents, quasi-rents; wages, employment levels, job security)
- (All of these factors can be viewed as outcomes as well as causes of course)

1.2 Labor standards and regulations

- Minimum wage laws
- Safety regulations
- Benefit mandates
- Overtime laws, workweek, hours per day/week
- Unemployment, injury compensation and disability insurance systems
- Dismissal and severance pay
- Tax treatment of various forms of income and investment (labor and capital taxation, investment depreciation, tax treatment of profits vs. revenue vs. value-added)

- Occupational licensing and certification

1.3 Social norms

- Norms of fairness and reciprocity
- Tolerance or aversion to inequality
- Identification with work, with firm, with social role

1.4 Competitive environment

- Market power of firms
- Presence of rents
- Degree of international competition

2 Unions and productivity

There is a huge body of literature on the union wage premium, on the efficiency of union bargaining, and (to a lesser extent), employer's strategic investment and financing responses to union activity. Much less studied, but even more intriguing—at least to me—is the effect of labor unions on worker and firm productivity. The classic book by Freeman and Medoff, *What Do Unions Do?*, argues that one of the key functions of labor unions is to give voice to worker's ideas, preferences, and private information in a way that is likely to be productivity enhancing. Thus, for Freeman and Medoff, the question was not whether unions raise productivity but whether the productivity benefits that they offer to firms are, from the firm's perspective, more than offset by the additional wages and benefits that they extract.

Five papers on your syllabus speak directly to the question of union's and productivity. The papers by Krueger and Mas (2003) and Mas (2008) study specifically how labor unrest—that is, a period of strike or strife activity—affect the quality of production. Both papers find that output quality is significantly degraded (in one case, in the manufacture of tires, in the other in the manufacture of construction equipment). These studies are clever and important, but at another level, one might argue that they're unsurprising. After all, strikes are *intended* to be disruptive. So, it's not entirely shocking that intricately produced goods built during strike periods suffer a bit in the quality dimension. Perhaps more to the point, strikes are *not* the normal state of affairs with labor unions. Even if labor unions improve productivity or quality on average, we won't learn about that by studying strikes.

The papers by Schmitz (2005), Mas (2006), and Lee and Mas (2012) provide something closer to an "average causal effect" estimate of the consequences of unionization for productivity. The Schmitz paper studies productivity of Great Lakes iron ore producers, which had faced no foreign competition in the Great Lakes steel market for nearly a century as of the late 1970s. In the early 1980s, as a result of unprecedented developments in the world steel market, Brazilian producers began delivering iron ore to Chicago (the heart of the Great Lakes market) at prices substantially below prices of local iron ore. Schmitz's paper offers a case study of the dramatic productivity increases in Great Lakes iron ore production that ensues, and traces these gains to changes in labor practices. Mas' 2006 paper examines how police unions' success or failure in binding offer arbitration cases affects their subsequent productivity. Finally, Lee and Mas (2012) study stock market reactions to unions' success or failure in representation elections held in private sector firms. While none of these papers directly establishes an "average" union productivity effect, they do paint a broadly consistent picture that suggests that even absent strikes, unions do in many cases adversely affect firm performance.

An even broader question that these papers raise is what precisely do unions maximize—that is, what are the arguments in a union's utility function? Schmitz's 2005 paper ends with a thought provoking (if somewhat crudely articulated) set of observations on this point:

> I have shown that increases in competition (or decreases in tariffs) led to surges in TFP through changes in restrictive work practices. This naturally leads to the question, Why were restrictive work practices not changed before the crisis in iron ore? And why do they persist today in nearly full force in Class I railroads?
>
> Let me start with a straw man. This straw man says that these work practices were part of a rent package received by workers. In this view, work practices led to idle time that was valued by the workers. In other words, workers used some of their rents to purchase idle time and other non-pecuniary benefits. With increased competition, rents were destroyed; hence, work practices had to be changed.
>
> But this view is vastly incomplete, if there is much truth to it at all. If it was idle time workers wanted, why structure work practices so that machinery sat idle as well? With machinery idle, capital productivity and materials productivity suffer. Work practices clearly led to money being flushed down the toilet. I cannot say this loud enough. Hence, there are other reasons these work practices were not changed before the crisis.
>
> What are these other reasons? I mention two possibilities. It is beyond the scope of this paper to do more. Money can obviously be made by changing such work practices. But there may be disagreements among groups (e.g., workers

> vs. local managers, repair workers vs. other workers) about how to divide the money. And, of course, there may be commitment problems. Can groups be assured that agreements will be honored? Also, might an "outside" group, such as the local towns, say, through increased taxation, attempt to capture some of the gains? Many steel companies today, in fact, are trying to dump pension obligations made to early retirees in the 1980s. Of course, whatever the reason work practices were not changed, with competition, it may no longer be possible to flush money down the toilet and survive (as was the case here), and restrictive work practices must be changed.

Stated formally, Schmitz' observation is that union rent extraction was not efficient; if simple rent extraction were the goal, there should have been less costly ways to extract the same amount of rents without producing so much ancillary wastage. There are at least two broad explantations for these facts. One is that we've misunderstood the nature of rents from the perspective of the union. Perhaps union members get direct utility from watching expensive equipment sit idle. Or, more plausibly, they direct value exercising *control rights* over capital, job assignments, and work processes. Thus, unions don't fight to establish arcane work rules simply to allow their members to slack off, but instead because workers value the autonomy and authority derived from their "rights" to do or not do certain work, to change shifts ritualistically in a designated shift-change location rather than "eyeball-to-eyeball" at the machinery, and to know with contractual certainty that the jobs that they've mastered will not be eliminated by technology or work reorganization.

An alternative explanation for these same facts is that the impossibility of striking credible, enforceable, time-consistent efficient bargaining leads to ancillary costs, like idled equipment, overstaffing, and inflexible work rules. However, if firms and unions could strike an efficient bargain, then both would strictly prefer to eliminate these forms of wastage and have the firm make a direct transfer to workers instead. The classic reference on this second viewpoint is Grout's 1984 *Econometrica* paper, "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," which shows how the possibility of ex post holdup constrains firms from making efficient up front investments—so much so that a union with too much bargaining power could end up extracting less rents than an identically situated union with less bargaining power. See also Carmichael and MacLeod, "Worker Cooperation and the Ratchet Effect" in the *Journal of Labor Economics* in 2000. Their paper makes the argument that competition among firms inevitably eliminates (makes non-enforceable) bargains in which workers are rewarded for implementing productivity improvements in a piece rate setting.

Personally, I find *both* of these viewpoints somewhat compelling. I'd be reluctant to say

that the explanation for the puzzle of "inefficient" union rent extraction is entirely due to one or the other.

3 Investment and bargaining in the absence of binding contracts

The Grout (1984) paper is a classic that makes a fundamentally important point. I suspect you'll find this point intuitive, but it's sufficiently important that it's worth reviewing. Here are the basics:

- A firm uses both capital, k, and labor, l, and has a profit function $\pi(k, l)$, which is strictly convex in input prices and twice differentiable with cross partials of constant sign.
- The purchase price of capital is c per unit, but once installed capital has a resale price of $q < c$.
- Shareholders are interested in maximizing profits and workers in maximizing the total income of the union, w. The size of the union is L. Any member of the union not employed by the firm receives w_0 elsewhere in the economy.
- Either group (firm or union) can prevent production from taking places. Thus, there is a potential hold up problem.

3.1 Efficient benchmark

- If contracts are binding, capital and labor will be employed efficiently. By Shephard's lemma:

$$k = -\pi_1(c, w_0),$$
$$l = -\pi_2(c, w_0).$$

[Recall Shephard's lemma: at the optimal input choice $e^*(p_1, p_2, ..., p_n)$, the derivative of the expenditure function with respect to the cost of any of the inputs is simply the demand for that input: $e_n^* = x_n$. This follows from the envelope theorem. At the interior maximization of the profit maximization function, the marginal revenue product of each input is exactly equal to its price. Since we are at a saddle point of the function, an epsilon change in the cost of an input has only a second order effect on the optimized choice of the *quantity* of that input. But of course, *expenditure* on the input will rise by epsilon times the initial quantity of that input being purchased. Thus $\partial e/\partial p_x = x$.]

- Now imagine that profits are non-zero in this setting. This could be because labor is restricted, for example ($L < l^*$). Alternatively, the union may have more members than there are jobs $L > l^*$. In that case, firm and union will need to bargain. Grout adopts the generalized Nash bargain, which gives the following maximand if $L > l^*$:

$$p^{\alpha} \left(\pi \left(c, w_0 \right) - p \right)^{1-\alpha},$$

  where p is the firm's retained profit and $\pi \left(c, w_0 \right) - p$ is the profit distributed to workers. If $L < l^*$, the maximand becomes:

$$p^{\alpha} \left(\pi \left(c, w_0, L \right) - p \right)^{1-\alpha},$$

  where $\pi \left(c, w_0, L \right)$ is the restricted profit function with labor restricted to $L < l^*$.

- The firm's profits will be given by $p = \alpha \pi \left(c, w_0 \right)$ or $p = \alpha \pi \left(c, w_0, L \right)$ in the latter case.

- Theorem 2.2 of Grout offers the key observation. If binding contracts are available, total profits are independent of the division of surplus between profits and union wages. Maximizing the economic pie and dividing the pie are separable problems, with the ultimate division of surplus having no impact on economic efficiency.

3.2 Binding contracts absent

- If contracts are absent, then the firm's ex ante sunk investment decision affects the nature of the bargain. In particular, the union may capture some share of the difference between the purchase and resale price of the firm's sunk capital investment, $(c - q) k$. In fact, the union cannot commit *not* to capture the quasi-rents on the firm's sunk investments. Recognizing this, the firm will alter its investment decision. This is a 'holdup' problem.

- To see the effect on investment, think of the holdup problem as changing the firm's effective price of capital. Specifically, call this cost γ (instead of c). So, the firm's choice of capital will satisfy $k' = -\pi_1 \left(\gamma, w_0 \right)$.

- We can continue to assume that labor is chosen efficiently since the opportunity cost of labor is w_o at all times, and there is no holdup problem.

- The union and the firm now bargain over the following surplus

$$p^{\alpha} \left(\pi \left(\gamma, w_0 \right) - \left(\gamma - q \right) \pi_1 \left(\gamma, w_0 \right) - p \right)^{1-\alpha},$$

where we are using the fact that the firm's threat point includes $-q\pi_1(c, w_0)$ since it could always resell its sunk investment at this price. Notice if $q = 0$, then the full cost of capital is subject to holdup.

- Profits of the firm are now given by

$$p' = \alpha(\pi(\gamma, w_0) - (\gamma - q)\pi_1(\gamma, w_0)) + (c - q)\pi_1(\gamma, w_0).$$

Differentiating with respect to γ and setting the FOC to zero gives

$$(\pi_1(\gamma, w_0) - \pi_1(\gamma, w_0)) + \pi_{11}(\gamma, w_0)(c - q - \alpha(\gamma - q)) = 0.$$

(Note that π_1 terms cancel as we apply the chain rule.) Since π_{11} will be non-zero due to the convexity of the profit function, we can simplify this FOC to:

$$c - q - \alpha(\gamma - q) = 0.$$

- Rearranging, we get the following expression for the firm's effective cost of capital in the absence of binding contracts:

$$\gamma = \frac{c}{\alpha} + \left(1 - \frac{1}{\alpha}\right) q.$$

Notice that if $\alpha = 1$, that is, the firm has all of the bargaining power, or $q = c$, so that capital can be resold at its purchase price (it is effectively not sunk), then $\gamma = c$, meaning that the absence of binding contracts has no effect on investment. If neither of these conditions holds, however, then bargaining has real effects on investment because it effectively increases the firm's capital cost ($\gamma > c$).

- From there, it's immediate that the absence of binding contracts generates inefficiently low investment except in the corner cases ($\gamma = c, \alpha = 1$). Profits are necessarily lower than they would be absent binding contracts. Moreover, an increase in wages, stemming from a rise in union bargaining power, reduces profits more than one for one.

- Total wages paid may *also* fall as α rises. Although the union gains a bigger piece of the economic pie, the pie is shrinking as the union's power increases.

- I find it's useful to have Grout's insights in mind when thinking about the Schmitz paper.

- The 2013 NBER paper by Greenwood and Weiss on "Mining Surplus Theory" applies a similar framework to the Schmitz results to offer a "structural" interpretation. I didn't find this paper too insightful, but I guess it gives some sense of why such bargaining can be inefficient.

4 Mas (2006): Reference points

One of the most enduring idea's of Akerlof's famous 1982 *QJE* gift exchange paper is *not* the specifics of making reciprocal gifts but rather the notion that individuals use a *reference point* when making judgments about the fairness of a transaction. This idea is also found in Kahneman and Tversky's epochal 1979 *Econometrica* article on Prospect Theory. Here, endowments serve as reference points, with the concavity of the utility function being discontinuously greater above the reference point than below—which gives rise to the odd, but central, prediction that willingness to pay is often substantially below willingness to accept, even for very small deviations in wealth.

The paper by Alex Mas on pay and reference puts these important ideas to excellent use. Mas' idea is to study how outcomes of Final Offer Arbitration (FOA) affect the productivity of police departments. FOA works as follows: in the event of a disagreement between workers and managers, each party submits a final offer and the arbitrator chooses between them (without modification). This mechanism creates a type of uncertainty that may be salutary. Each party has the incentive to moderate its offer to increase the odds that it is chosen by the arbitrator. Stevens (1966) [as quoted in Farber, 1980] has argued that FOA "generates just the kind of uncertainty... that is well calculated... to compel them [the parties] to seek security in agreement." The insight of the Mas paper is that this form of arbitration potentially yields a 'reference point' in the form of the union's offer against which bargaining outcomes will be judged. If worker effort is indeed based on a 'reference wage' as in Akerlof 1982, deviations from the union's proposed FOA bargain may cause workers to reduce effort. Certainly, this is a creative hypothesis. [It's worth asking, however, how interesting the paper would have been if the hypothesis had been rejected.]

4.1 Conceptual framework

The paper does not offer a model of reference dependent preferences (unlike Akerlof, 1982). It does review the basic arbitration model, which is useful background information for any labor economist (for details, see Farber, 1980, in the *Journal of Conflict Resolution*). In the basic arbitration model, it is assumed that workers, firms, and arbitrators observe a common value of the expected productivity of the workforce. The purpose of the arbitration is to

choose a division of the surplus. Arbitrator's choose a value Y_f, which is their view of the 'fair' allocation of the surplus. Firms and workers must each propose an allocation. The assumption is that the arbitrator chooses whichever is closer to his preferences. Thus, the arbitrator chooses the employer's proposal iff

$$|Y_e - Y_f| \leq |Y_u - Y_f| .$$

Assume that employers and unions have a common prior on the distribution of Y_f, but that there is uncertainty on the exact value of this Y_f. In making their proposals, employers and unions recognize that their offers trade off between the chances of winning the arbitration and the expected surplus conditional on winning. More risk averse parties will therefore make more conservative (closer to Y_f) offers.

Though not demonstrated in the paper, Mas notes on the authority of Farber (1980) that the following results should hold in equilibrium:

1. If both parties are equally risk averse, the winner in arbitration is determined by a coin toss. The intuition for this is clear: with common beliefs about the distribution of Y_f and identical utility functions, both parties will choose offers that equally trade off the odds of winning against the gains conditional on a win. Thus, they should face even odds of victory.

2. If parties have constant absolute risk aversion (CARA), the more risk averse party will be more likely to win. The intuition here is that the more risk averse party submits a 'more reasonable' offer. The reasons for assuming CARA is that it makes the utility functions comparable. We can't compare behavior of more general utility functions on the basis of a single parameter like the coefficient of absolute risk aversion. [I do not understand Mas' claim that the probability of an employer win is fixed, and therefore invariant to the facts of the case. I don't see this in Farber 1980 either, though I doubt the claim is incorrect.]

3. The offer spread $Y_e - Y_u$ is increasing in the uncertainty regarding the arbitrator's preferred award. Clearly, if Y_f were known with certainty, both employer and union would bid arbitrarily close to this value rather than risk losing with near certainty. Thus, it is uncertainty regarding the arbitrator's preferences that drives the equilibrium divergence in employer and union offers. (With symmetric utility f'ns, priors, and uncertainty, the employer and union will diverge to equal and opposite degrees from their prior on the arbitrator's median preference.)

The key identification condition for the Mas paper is that the arbitration outcome must not

be correlated with past police performance (otherwise, it would not be an exogenous shock to police performance). Under this exclusion restriction, the arbitrator's decision can be legitimately viewed as a shock to future earnings, potentially relative to some reference point (i.e., the union's offer). This surprise that provides the 'experiment' that the paper analyzes.

4.2 Measurement

The main policing outcome that is uniformly available is the number of crimes cleared per arrest per 100K residents in a municipality. Clearances refer to the number of crimes "solved" by the arrest of one or more persons. Arrests represent costly effort for police. Of course, more arrests are not always better (if police arrest the wrong people to fill quotas, for example). We will hope that this is not the case in the relevant range of arrest activity.

The FOA data come from New Jersey for years 1978 and 1995, and include offers submitted to the arbitrator (as a percentage of the prior wage) and information on whether the arbitrator ruled in favor of the municipality or the union. There are 383 arbitration cases from 255 cities.

Crime data are monthly counts from the FBI Uniform Crime Reports (UCR) for 1976-1996.

4.3 Results

- Table II shows that unions win 65 percent of cases, which may suggest that unions are more risk averse than employers.
- Importantly, there are no obvious differences in crime and clearance rates in the prior 12 months between municipalities in which unions win and those in which they lose.
- Figure I shows that following arbitration, clearances per capita diverge between cities in which unions win and those in which they lose. This figure is generated using only data for cities in which arbitration occurred. Is that problematic?
- Consider the following diff-in-diff equation:

$$y_{jt} = \alpha + \beta_1 \times UnionWin_j \times Post_t + \beta_2 \times UnionLose_j \times Post_t + \delta_j + \gamma_t + e_{jt},$$

  where δ and γ are a complete set of city and time effects (thus, the $\beta's$ are identified by the time $\times$ city interaction). Note that although in every case, there is one winner and one loser, the arbitrator decides which party takes which role. A potential issue here is that both winners and losers are treated, meaning that there are really two treatments.

We may theoretically be interested only in the relative effect (the difference between winning and losing). We could could still identify this relative effect if all arbitration cases were decided simultaneously; the time effect would be removed in contrasting the change in clearances for cities where the union lost versus won. But we are interested in the asymmetric effect of winning versus losing, so we really need to estimate both β_1 and β_2. This is feasible here sine cities arbitrate at different points in time. Cities that do not arbitrate in period t serve as a control for those that do. We can thus estimate the Union-Wins contrast in arbitrating relative to non-arbitrating cities and the Union-Loses contrast in arbitrating versus non-arbitrating cities. This is essentially what Figure I accomplishes.

- To potentially get a better estimate, the next figure uses non-arbitrating counties as a control group:

$$y_{t\tau bc} = \alpha + \psi_{_{FE}} + \beta_{1\tau} \times UnionWins_b + \beta_{2\tau} \times EmployerWins_b + \varepsilon_{t\tau bc},$$

where t is calendar time, τ is 'event time' (time relative to arbitration), and b denotes the arbitration window, meaning the subset of the interval after which the arbitration is decided. The fixed effects in $\psi_{_{FE}}$ include a set of arbitration window effects γ_b (one for each treated city during the arbitration window), a set of month by year fixed effects, and a set of city fixed effects. Coefficients $\beta_{1\tau}$ and $\beta_{2\tau}$ cannot be identified for each specific time interval in b because of the inclusion of γ_b. Thus, $\beta_{1\tau}$ and $\beta_{2\tau}$ are identified relative to $\tau = 0$. Figure II plots $\hat{\beta}_{1\tau}$ and $\hat{\beta}_{2\tau}$ for $\tau = \{-23, ..., -1, 1, ..., 23\}$. The fact that Figure II looks a lot like Figure I indicates that the winning and losing cities are highly comparable without regression adjustment.

- Table III. It is somewhat reassuring (as both a validity-test and policy matter) that arbitration outcomes do not appear to affect murder or rape clearances. They mostly affect assault, robbery and larceny clearances (which are the most commonplace—so also the easiest to identify).

- Table IV. Some limited evidence that crime rises when a union loses. If a union loss raises the crime rate, it should mechanically tend to raise the clearance rate (since clearances are per capita not per crime, so more crime allows higher clearance per capita). The crime results therefore reinforce the conclusion that union losses adversely affect police performance.

- Table V is a bit disturbing for the plausibility of the estimates: a union win leads to a 22% increase in the probability of incarceration conditional on the charges levied.

Conditional on conviction, the probability of incarceration rises by 25 percent and the sentence length rises by 25%. These are very large effects.

- Note that the average spread in an arbitration ruling is only 1.5% of pay!
- Figure V strongly suggests that reference points matter. The productivity effects are discontinuous at a gain of zero. Note that award minus average offer is suppose to proxy for the deviation of the award from the expected award (assuming the arbitrator flips a coin).
- Table VI presents a variety of evidence. The size of the loss matters when Unions lose, but seemingly the mere fact of winning is all that matters when unions win. Is this because losses hurt more than gains help?
- Another way to calculate expectations is to note that when the average of offers is higher, unions should be more likely to lose (this is in fact visible in Table I). The reason being that a higher average will tend to indicate that the employer is being more generous and the union more greedy. If the arbitrator's bliss point is normally distributed, then the probability of an employer win, p_b, can be modeled as a probit function where the explanatory variable is the average offer. Following this logic, Mas calculates:

$$E\left(award_b\right) = \hat{p}_b \cdot \text{Emp\_Offer}_b + \left(1 - \hat{p}_b\right) \text{Union\_Offer}_b.$$

  The expectations based award is then:

$$award_b - E\left(award_b\right).$$

  Amazingly, this proves to be a better explanatory variable than the actual award in column (5) of Table VI (the actual award and the expectations based award are very highly correlated, however). Thus, the effect of a loss is greater when the gap between the union and city offers is larger and when the expectation that the union will win is higher.
- The effect of a union loss on police productivity lasts for over a year. We might expect faster fade-out if this were an individual level treatment. However, there may be a 'social multiplier' operative.
- Also note the use of the phrase 'hedonic treadmill'—an excellent nugget for impressing family members or persons of romantic interest.

- Do cities understand that unions react badly to losses and therefore make submit offers that are unlikely to be selected? (Recall that unions win two-thirds of arbitrations.)
- See also Koszegi and Rabin (2006) for an intellectually attractive model of reference dependent preferences.

4.4 Conclusions

This is a nice example where a 'behavioral anomaly' is shown to have a substantial, durable impact on a consequential societal outcome. One interesting observation is that there is not obviously a market mechanism to arbitrage the reference point effect on police performance. If preferences for buying or selling stocks (or houses) were reference dependent, it's likely that the market would be able to better arbitrage these preferences. An interesting research agenda is to assess whether reference dependence is a broadly important behavioral phenomenon across numerous domains.

I view this paper as affirming a key tenet of the Akerlof (1982) model—reference points are a benchmark upon which (some) economic behavior depends. It's possible that the phenomenon of reference dependence is an even more enduring contribution of Akerlof's 1982 article than the gift exchange equilibrium itself. It's logical that Mas thanks George Akerlof in the paper's acknowledgments.

14.662 Labor Economics II

Spring 2015