Lecture Note 7: Fairness and Norms

David Autor
MIT 14.663 Spring 2009

May 31, 2009

1 Fairness and Norms

"...The phrase [fairness] is constantly used in the market place; it is frequent in the mouths both of employers and the employed; and almost every phrase in common use has a real meaning, though it may be difficult to get at." – Alfred Marshall

The starting point of models of fairness in economics is the Akerlof 1982 article on gift exchange. The model is worth understanding. And it has sparked a considerable literature, which is currently quite active.

2 Akerlof gift exchange model (1982)

2.1 Fair wage

Fair wage is:

$$w_{i,t+1}^{f} = f\left(w_{it}, w_0, b_u, u, e_i, e_0\right),$$

where

- $w_{i,t+1}^{f}$ is the perceived wage of i at $t+1$
- $w_{i,t}$ is the actual wage paid to i in t
- w_0 is the wage paid to others in $i's$ reference set
- b_u is the unemployment benefits of the individual's reference set in the current and previous periods
- u is the number of unemployed in the reference set in the current and previous periods
- e_i is the individual's work rules in current and previous periods
- e_0 is the work rules of people in the individual's reference set in current and previous periods

2.2 Norms

Norms equation:

$$e_n = e_n\left(\left\{w\left(e, \epsilon\right)\right\}, e_{\min}, u_1, ..., u_J; w_0, u, b_u\right),$$

where

- $\{w(e,\epsilon)\}$ is the function that relates wages of a worker of type ϵ to his effort (thus, this is the firm's remuneration system).
- $e_{\min}$ is the work rule
- u_j is the utility of the j_{th} worker in the firm
- w_0 is the wage paid by other firms
- u is the unemployment rate
- b_u is the unemployment benefit

2.3 Workers

A worker who has been offered a job must decide on his effort level and whether or not to accept the job. Utility depends on norms for effort, effort exerted, wage rate if employed, unemployment benefit if unemployed. There is a distribution of worker tastes $f(\epsilon)$.

Worker makes two choices: (1) if offered a job, must decide whether or not to accept the offer (exchange gifts); (2) if the offer is accepted, must decide the size of the reciprocal gift.

Worker of type ϵ has the utility of working for the firm of:

$$u(e_n, e, w, \epsilon),$$

and if not working for the firm:

$$u(b_u, \epsilon).$$

If working for the firm, the worker chooses the level of effort e which maximizes utility subject to effort meeting the firm's minimum job requirement (this may not bind). So, worker takes a job iff:

$$\max_{e \geq e_{\min}} u(e_n, e, w, \epsilon) > u(b_u, \epsilon).$$

2.4 Firms

Output is:

$$q = f\left(e_1, e_2, ...e_J\right),$$

where J is the number of workers hired.

The firm chooses the wage function $w\left(e, \epsilon\right)$, work rules $e_{\min}$, and number of workers to maximize:

$$pf\left(e_1, ...e_J\right) - \sum_{j=1}^{J} w\left(e, \epsilon\right),$$

where p is the output price. It is assumed that workers of type ϵ are offered jobs at random.

2.5 Model 1: Equilibrium unemployment

Let $\bar{l}$ workers per firm be the supply of labor. That is, this is the number of workers divided by the number of firms.

Let output be:

$$q\left(e, n\right) = \left(en\right)^{\alpha}.$$

There is no worker heterogeneity, and all workers will exert effort equal to the norm:

$$e = e_n.$$

Let the effort norm be a function of the firm's wage relative to the reference wage:

$$e_n = -a + b\left(w/w_r\right)^{\gamma}, \ \gamma < 1.$$

Let the reference wage be the geometric mean of the outside wage and the unemployment benefit:

$$w_r = w_0^{1-u} b_u^u,$$

where u is the unemployment rate, w_0 is the wage paid by other firms, and b_u is the unemployment benefit. So

$$e_n = -a + b\left(w/w_0^{1-u} b_u^u\right)^{\gamma}, \ \gamma < 1$$

Assuming that this firm is the typical firm:

$$n = \left(1 - u\right)\bar{l},$$

and

$$w = w_0.$$

That is, it has the typical employment and wage levels.

Suppose that $u = u_0 > 0$. If $u_0 > 0$, the paper asserts that the firm can obtain all the workers it wants at any wage. [I don't see where the reservation wage $u(b_u, \epsilon)$ is entering here, but let's assume that this constraint is non-binding.] If so, the firm faces the maximization problem:

$$\begin{aligned}
\max_{n,w} \Pi &= (en)^{\alpha} - wn \text{ s.t.} \\
e &= e_n \\
&\quad e_{n=} - a + b\left(w/w_r\right)^{\gamma} \\
w_r &= w_0^{1-u} b_u^u.
\end{aligned}$$

Plugging in:

$$\begin{aligned}
\frac{\partial \Pi}{\partial n} &= 0 = \alpha e^{\alpha} n^{\alpha-1} - w \\
\alpha e^{\alpha} n^{\alpha-1} &= w \\
n^{\alpha-1} &= \alpha^{-1} w e^{-\alpha} \\
n &= \left[\frac{w}{\alpha e^a}\right]^{1/(\alpha-1)} \\
n^* &= \left[\frac{w}{\alpha\left[-a + b\left(w/w_r\right)^{\gamma}\right]^a}\right]^{1/(\alpha-1)}.
\end{aligned}$$

One useful step here (though this can be collapsed into the steps above) is solve for the 'Solow condition' in the efficiency wage model. The Solow condition is the observation that if effort depends on the wage, then at the optimal wage level, the elasticity of effort with respect to the wage must be one (otherwise, the wage is too high or too low).

Solow condition:

$$e_n = -a + b\left(w/w_r\right)^{\gamma}$$

$$\begin{aligned}
\frac{\partial e}{\partial w} &= \gamma b \left(w/w_r\right)^{\gamma-1} w_r^{-1} \\
&= \gamma b w^{\gamma-1} w_r^{-\gamma}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial e}{\partial w} \times \frac{w}{e} &= 1 = \frac{w\gamma b w^{\gamma-1} w_r^{-\gamma}}{-a + b\left(w/w_r\right)^{\gamma}} \\
-a + b\left(w/w_r\right)^{\gamma} &= w\gamma b w^{\gamma-1} w_r^{-\gamma} \\
b\left(1-\gamma\right) w^{\gamma} w_r^{-\gamma} &= a \\
w^{\gamma} w_r^{-\gamma} &= a/\left(b\left(1-\gamma\right)\right) \\
w^{\gamma} w^{(u-1)\gamma} b_u^{-\gamma u} &= a/\left(b\left(1-\gamma\right)\right) \\
w^{(u-1)\gamma+\gamma} &= b_u^{\gamma u} a/\left(b\left(1-\gamma\right)\right) \\
w^{u\gamma} &= b_u^{\gamma u} a/\left(b\left(1-\gamma\right)\right) \\
w &= b_u \left[a/\left(b\left(1-\gamma\right)\right)\right]^{1/u\gamma} \\
w &= b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma}.
\end{aligned}
$$

Notice that $\partial w/\partial u < 0$. Thus, the Solow wage is falling in the unemployment rate. But notice also that $w > w_r$ under the following condition:

$$
\begin{aligned}
b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} &> \left[b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma}\right]^{1-u} b_u^u \\
b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} &> b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{(1-u)/u\gamma} \\
\left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right] &> \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{(1-u)} \\
1 &> \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{-u} \\
1 &> \frac{a}{b\left(1-\gamma\right)},
\end{aligned}
$$

which means that the Solow wage is greater than the reference wage iff the elasticity of the effort with respect to the wage is initially greater than 1, meaning that increases in the wage beyond the reference wage pay for themselves.

Using this value, we can plug back into the maximization:

$$\begin{aligned}
n^* &= \left[\alpha^{-1} w \left(-a + b\left(w/w_r\right)^{\gamma}\right)^{1-\alpha}\right]^{1/(\alpha-1)} \\
&= \left[\alpha^{-1} b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} \left(-a + b\left(w/w_r\right)^{\gamma}\right)^{1-\alpha}\right]^{1/(\alpha-1)} \\
&= \left[\alpha^{-1} b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} \left(-a + b w^{u\gamma} b_u^{-u\gamma}\right)^{1-\alpha}\right]^{1/(\alpha-1)} \\
&= \left[\alpha^{-1} b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} \left(-a + b\left(b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma}\right)^{u\gamma} b_u^{-u\gamma}\right)^{1-\alpha}\right]^{1/(\alpha-1)} \\
&= \left[\alpha^{-1} b_u \left[\frac{a}{\left(b\left(1-\gamma\right)\right)}\right]^{1/u\gamma} \left(-a + \frac{a}{\left(1-\gamma\right)}\right)^{1-\alpha}\right]^{1/(\alpha-1)} \\
&= \left[\alpha^{-1} b_u \left(\frac{a\gamma}{1-\gamma}\right)^{1-\alpha} \left(\frac{a}{b\left(1-\gamma\right)}\right)^{1/\gamma u}\right]^{1/(\alpha-1)}.
\end{aligned}$$

If this demand function is consistent with the unemployment rate, then the supply of labor must be:

$$\bar{l} = l_0 = \frac{n^*}{1-u} = (1-u)^{-1}\left[\alpha^{-1} b_u \left(\frac{a\gamma}{1-\gamma}\right)^{1-\alpha} \left(\frac{a}{b\left(1-\gamma\right)}\right)^{1/\gamma u}\right]^{1/(\alpha-1)}.$$

What do we learn from this model?

1. The first point is that if worker effort depends on a 'reference wage,' then it may be logical for firms to pay a wage above that level to obtain extra effort (under the condition spelled out above).

2. If so, the profit maximizing choice of labor input will not equate the marginal product with the reservation wage but rather with the 'Solow wage' (AKA, the efficiency wage).

3. At this wage level, there will be equilibrium unemployment. That is, workers will be willing to work at w, but firms will have no incentive to hire them. Why not? Because the marginal product of labor at n^* is equated with the efficiency wage (Solow wage). Further hiring would lower the marginal revenue product of labor, so the firm would have an incentive to cut the wage. But cutting the wage would reduce effort of all of its

workers, thus lowering their productivity. The firm does not therefore face an incentive to deviate from w given the market equilibrium at u.

4. Since the reference wage (as parameterized here) is falling in the unemployment rate (and hence effort is rising in the unemployment rate), wages will fall and labor demand will rise with increases in unemployment. Notice that

$$\begin{aligned}
\ln n &= \ln\left[\alpha^{-1} b_u \left(\frac{a\gamma}{1-\gamma}\right)^{1-\alpha} \left(\frac{a}{b\,(1-\gamma)}\right)^{1/\gamma u}\right]^{1/(\alpha-1)} \\
\ln n &= \left[1/\left(\alpha-1\right)\right]\left[1/\gamma u \ln\left(\frac{a}{b\,(1-\gamma)}\right) + \ln\left[\alpha^{-1} b_u \left(\frac{a\gamma}{1-\gamma}\right)^{1-\alpha}\right]\right] \\
\ln n &= \left[\frac{1}{(\alpha-1)\,\gamma u}\right]\ln\left(\frac{a}{b\,(1-\gamma)}\right) + \left[1/\left(\alpha-1\right)\right]\left[\ln\left[\alpha^{-1} b_u \left(\frac{a\gamma}{1-\gamma}\right)^{1-\alpha}\right]\right] \\
\frac{\partial \ln n}{\partial u} &= \frac{1}{(1-\alpha)\,\gamma u^2}\ln\left(\frac{a}{b\,(1-\gamma)}\right),
\end{aligned}$$

   which is positive if (as above $a/b\,(1-\gamma) > 1$) and $\alpha < 1$. Clearly, it must be the case that $\alpha < 1$ or else the profit function would have increasing returns over its entire range and the firm would have unbounded demand for labor.

5. But also notice that if $b = 0$, then effort is not increasing in the wage. If $a < 0$, then $e = -a > 0$, so effort is positive at any wage. In this case, the firm will choose labor:

$$\begin{aligned}
w &= \alpha\left((-a)\,n\right)^{\alpha-1} \\
\alpha^{-1} w\,(-a)^{1-\alpha} &= n^{\alpha-1} \\
n &= \left[\alpha^{-1} w\,(-a)^{1-\alpha}\right]^{1/(\alpha-1)},
\end{aligned}$$

   which does not depend on the unemployment rate. And of course, the wage is equal to the marginal product of labor. So, we should expect that labor markets will clear (that is, all workers who want work at wage w will receive it—though of course, if many additional workers wanted work, this would lower the wage, but that's still market-clearing).

In summary, this model posits a labor market equilibrium where, because workers work harder when paid above their opportunity cost, firms pay above market clearing wages and

equilibrium unemployment results. Conceptually, this model is not very dissimilar from the seminal Shapiro-Stiglitz (1984) efficiency wage model, published two years later in the *AER*. The key difference vis-a-vis the SS model is that here, workers respond positively to higher pay, whereas in SS, workers are inhibited from shirking by higher pay. In SS, if a firm deviated by paying slightly more by paying above the market equilibrium, it would not be rewarded with higher effort. In the Akerlof model, higher pay always leads to higher effort (though firm's clearly would not wish to pay for unbounded effort).

I will not go through the second Akerlof model, in which work standards influence worker behavior.

3 Evidence on gift exchange

There is a contentious literature on gift exchange, and I would say that there is no definitive paper on the topic. An early and influential paper is by Fehr, Kirschsteiger and Riedl (1993), who provide a laboratory test of gift exchange. I will discuss this paper, as well as recent work by List (2006, *Econometrica*) that presents an opposing viewpoint.

3.1 Fehr, Kirschsteiger and Riedl (1993)

This lab experiment provides a template that countless subsequent studies have followed. The general ingredients are: (1) a contracting phase, usually anonymous, in which workers and employers agree on wages; and (2) a work phase, in which workers choose an 'effort' level, which, if positive reduces their payoff but increases the payoff of the employer.

Here's how it works:

- There are always more workers than employers (excess labor supply). The ratio is either 9:6 or 8:5.
- There is a three minute oral auction in which employers make blind wage offers to the pool of workers. Wage bids are in increments of five. The first worker to accept the offer gets the job. Only one worker can work for a given firm.
- Given this auction structure, workers have an incentive *not* to hold out for high wages.

If a wage offer is not accepted, an employer can bid again at a higher wage.

- After three minutes, the proceed to a work phases. Workers must pay a fixed cost $c = 26$ to work, and this is compulsory if they have accepted a contract.
- In addition, workers must choose effort according to the convex cost schedule in Table 1. The workers effort ranges from 0.1 to 1, and her cost of effort ranges from 0 to 18.
- The worker's payoff is therefore $u_i = w_i - m(e_i) - 26$, where w is the agreed wage end $m(\cdot)$ is the effort cost function.
- The firm's payoff is $\Pi_i = (126 - w_i) e_i$. Note that the firm cannot make negative profits with this payoff function.

3.1.1 What's supposed to happen?

Clearly, workers have no pecuniary incentive to exert effort beyond the minimum (remember that there is excess labor supply and an anonymous market). If firms expect that workers will exert $e_{\min}$ at any wage level, they should simply pay workers their reservation wage of 30 (remember, increments of 5). Thus, expected worker utility is:

$$E(u) = 30 - 0 - 26 = 4$$

and expected firm profits are:

$$E(\Pi) = (126 - 30)\, 0.1 = 9.6.$$

Clearly, this is not a welfare maximizing outcome. If workers were to choose maximal effort, output rises by a factor of 10 while workers cost rise by 18. Firms could easily compensate workers to make this effort and still be much better off $(126 - 45) = 81)$. Hence, if firms believed workers would exert greater effort if offered a higher wage, it would clearly be optimal for them to pay higher wages.

3.1.2 What happened?

It's pretty evident that firms paid above reservation wages and workers supplied above minimal effort. One interesting exercise. Define the overpayment as:

$$r = (w - 30) / (126 - 26),$$

which is simply the actual wage minus the reservation wage scaled by the surplus. At the competitive baseline, this should be zero. In point of fact, it's typically about 0.4, and is also highly correlated with the effort level exerted *after* the wage is set. There is no tendency for the market to converge towards the subgame perfect Nash prediction. And effort in the last period is always higher than in the first period.

It's also interesting to note that workers did not 'hold out' for high wages. In only 10 of 144 cases in which workers were unemployed was it because they rejected a wage offer. In the other 134 cases, it's because they were beaten out by other workers (who jumped on the wage bid). It bears note that there was no market-clearing wage possible in this game since there was excess labor available by construction.

Average profits per period were 18.9, relative to a predicted level of 9.6. It appears that average wages were about 80 (details are not given).

3.1.3 What explains the results?

Reputation formation was not possible—so this is not an explanation for the pattern of results.

Could it be that buyers and sellers wanted to seem fair?

It seems likely, as the authors note, that extreme rent asymmetries coupled with complete information about payoffs made it easy for equity considerations to enter the environment. This is a very non-standard work environment (even by the standards of 'workplaces' in which no one works) in that employers and workers know one another's payoffs exactly.

Conclusions: These are intriguing results. But one should be extremely cautious about the external validity of lab experiments. There are a number of well-known cases where such results prove highly non-robust to small deviations, suggesting that contextual effects are often surprisingly important. The Lazear-Malmendier-Weber 'sorting' critique also applies (I will

discuss this in class).

3.2 Gneezy and List, 2006

This paper makes a simple but controversial contribution to the Gift Exchange literature by probing the external validity of gift exchange experiments. Almost all of the evidence for gift exchange (perhaps *all* of the evidence) is based on either anecdotes or lab tests like in the FKR paper. Gneezy and List provide a sort-of test of gift exchange in the field.

3.3 Protocol 1: Data entry

Students were recruited to work computerizing the holdings of a library. Posters offered $12 an hour for *one-time* work, implying that this would not be a repeated interaction. Each participant performed the task alone without viewing other members. In the control treatment, participants were paid $12. In the experimental treatment, they were told after the training was complete that they would receive $20 per hour. Output was monitored every 90 minutes.

Students were not told that they were participating in an experiment. Nineteen workers were hired for 6 hours. There were 10 controls and 9 treatment subjects.

Results of this treatment are clear cut. In the first 90 minutes, treated subjects produce about 25 percent more output per hour. In the second 90 minutes, they produce 10 percent more output. In the next two 90 minute periods, their output is almost identical to that of untreated subjects.

Note that only 19 subjects were hired, and 10 of 19 were 'gifted.'

3.4 Protocol 2: Fund-raising

The task is door-to-door fund-raising for the Natural Hazards Mitigation Research Center. 23 fund-raisers were recruited, 13 in the no-gift condition and 10 in the gift condition. All fund-raisers were told they would be paid $10 per hour during training and employment. After the training was completed, subjects met on a Saturday morning to begin fund-raising. The gift-treatment subjects were told at that time that they would be receiving $20 per hour instead of $10.

Subjects were not told that they were participating in an experiment. The gift and no-gift groups were kept separate so that there was no possibility of spillovers/contamination. Subjects were instructed to write down the amount of each contribution and time it was received so that it was possible to track output over the course of the day.

Results are again clear. In the first three hours, the gift group raises about 75% more per hour than the non-gift group: \$11.00 vs. \$6.40 per hour. In the second three hours, the gift group raises only 5% more per hour than the non-gift group: \$7.03 vs. \$6.63. Clearly, paying twice as much per hour was not a worthwhile investment.

Only 23 subjects were hired, and 13 of 23 were 'gifted.'

3.5 Interpretations

An obvious conclusion here is that time span matters. Gift exchange did work initially—it's just that its effect was not durable. The authors offer one explanation for this: 'hot' and 'cold' responses. The initial 'hot' response is for participants to feel an upwelling of gratitude that increases motivation. The slightly longer term 'cold' response is for participant motivation to fall back to the 'standard' level, whatever that is. Why does this occur? Several possibilities:

- After the 'hot' response wears off, participants realize that there is no *marginal* incentive to work harder; no matter how little or much they fund raise, their pay is still \$20 per hour.

- One can also invoke Prospect Theory. When the reference pay is \$10, the \$20 rate leaves considerable psychic surplus. After a few hours, the reference rises to \$20 (participants now expect it) and so participants do not feel (much) more satisfied with their roles than they would have under the \$10 pay scheme.

- It would be interesting to apply a subsequent treatment in which pay was *reduced* to \$10 on a second day of fund-raising. I suspect that this would have *durable* effects—treated participants would produce less output than untreated participants during both halves of the day.

Does this experiment spell the death of the theory of Gift Exchange? I very much doubt it. One can lodge several objections about external validity:

1. The fact that this is a one-shot interaction seems at odds with the general gift exchange model where we believe that employers and employees continually 'regift' each other. Of course, one could respond that this setting has an alternative explanation: in repeated games without a finite (or known) duration, cooperation can emerge as a Nash equilibrium strategy (that is, the benefits over the long-term of cooperating outweigh the short-term gains that come from not returning the gift).

2. It may be that gift exchange is more salient for negative than positive reciprocity. As in my conjecture above, removing the gift may do more harm than adding the gift yielded benefits.

3. One can also invoke the LMW sorting critique. The subjects recruited for this study were not selected for reciprocity. If an employer is going to use a gift exchange wage policy, it will want to select workers who tend to respond cooperatively. Hence, gift exchange that operates within competitive markets might only take place with selected participants. Case in point, it's unlikely that an employer would, for no announced reason, double the bargained wage prior to work beginning. More artificial still in this case is that the employer does not announce that along with a doubling of the wage comes an expectation (or request or stated desire) for higher output. Thus, it's not entirely clear what the gift is supposed to stimulate in this setting. That said, output did rise initially, so there clearly is a reciprocation reflex, though it does not seem to be enduring.

3.6 More on external validity of lab experiments: List (2007), "On the Interpretation of Giving in Dictator Games."

This paper is almost too tongue-in-cheek to belong in a serious economics journal, but the substantive point made by the paper is still relevant: the 'cues' that cause participants to share in the dictator game (and other laboratory settings that appear to generate non-competitive behavior) may be highly non-robust and context-specific.

Baseline and three treatments:

1. Baseline: A one-shot, anonymous dictator game with \$5 to divide with choices on the range $[\$0, \$5]$ in 50 cent increments.

2. Treatment 1: the action range is extended to $[-\$1, \$5]$. That is, the dictators can take up to \$1 from the paired participant.

3. Treatment 2: symmetric action range $[-\$5, \$5]$.

4. Treatment 3: Same as Treatment 2 except that all agents first engaged in a 30 minute task for which they the dictators earned \$10 and the non-dictators received \$5. Subjects knew about one another's earnings. Not clear if they knew that they had received unequal pay for equal work.

As the figures make clear, this modest reframing has huge consequences on the frequency and amount of giving. Notably, in Treatment 3, in which dictators may believe that non-dictators *worked* for their \$5, there appears to be considerably more reluctance to take \$5.

3.7 Mas (2006): Reference points

One criticism of all of the work on gift exchange that we discuss is that it takes the idea extremely literally. To my mind, the most enduring idea in the Akerlof 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. The paper by Alex Mas on pay and reference puts this idea to excellent use. Moreover, while the experimental papers above are all more or less 'rigged' (IMHO), the Mas article on pay and reference points is decidedly real-world.

The idea of this paper 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. Accordingly, if worker effort is indeed based on a 'reference wage' (as in Akerlof 1984), deviations from the union's proposed FOA bargain may cause workers to reduce effort. Certainly, this is a creative and plausible hypothesis. [It's worth asking, however, how interesting the paper would have been if the hypothesis had been rejected. If the answer is 'not very interesting,' this suggests that this is a research project whose success is result-dependent, which is not altogether good.]

3.7.1 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 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.

The key identification condition for the Mas paper is that the arbitration outcome must not be correlated with past performance. If so, it can be legitimately viewed as a surprise adjustment to earnings, potentially relative to some reference point (i.e., the union's offer). It is this surprise that provides the 'experiment' that the paper analyzes.

3.7.2 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.

3.7.3 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:

$$y_{jt} = \alpha + \beta \times Treat_j \times Post_t + \delta_j + \gamma_t + e_{jt},$$

where δ and γ are a complete set of city and time effects (thus, β is identified by the time $\times$ city interaction). A potential issue here is that both winners and losers are treated, meaning that there are really two treatments. We may be interested only in the relative effect (the difference between winning and losing), but this effect would *not* be identified if all arbitration cases were decided simultaneously. The reason is that both groups (winners and losers) would be treated at one time—meaning there is no reference group that serves as a contrast for the treated group. However, the model would still be estimable if the econometrician designated only one group as treated (let's say losers), and so the winners would serve as the reference group (thus $Treat_j$ would be replaced by $Lose_j$). We would get a contrast, but it would not identify a parameter of interest. It should be clear that it is not possible to identify both the treatment and time effect when all groups are treated simultaneously.

In the case of the Mas estimates, however, this issue is not present because 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 winner/loser contrast in arbitrating relative to non-arbitrating cities. This same strategy can also be used to estimate both treatment effects (the effect of winning and the effect of losing), and this is essentially what Figure I accomplishes. Note that although in every case, there is one winner and one loser, the arbitrator decides which party takes each role.

- One minor threat to validity: what if there is simply an 'arbitration effect' on outcomes that exists independently of the arbitrator's choice. In this case, we cannot separately identify the winner and loser effect. But we can identify the relative effect.

- 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 identify 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 most 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 unions lose. 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, indicating a 22% increase in the probability of incarceration conditional on the charges levied if the union wins. *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 closer to the arbitrator's preferred point and the union is further away. 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:

$$E\left(award_b\right) = \hat{p}_b \cdot EmpOffer_b + \left(1 - \hat{p}_b\right) UnionOffer_b.$$

The expectations based award is then:

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

Amazingly, this proves 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 police loss effect lasts for over a year, which is far more than the three hour gift exchange effect found by Gneezy and List (2007). 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 targets 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.

3.7.4 Conclusions

This is the best example I know of in which a 'behavioral anomaly' is found 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 important 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. If Akerlof were *not* alive right now, he would be smiling down from the afterworld on Alex Mas' paper. Seeing as Akerlof is currently alive, it's logical that Mas thanks him in the paper's acknowledgments.

4 Other evidences on norms and beliefs

Most economic research that dabbles the area of 'fairness' and norms attempts to ask whether norms and beliefs cause agents to differ from *homo economicus*. But we really don't much about how norms and beliefs are formed. I'll discuss two papers that have something to say about this. The paper by Gneezy in 2005 has a few nice insights into the contours of agents' moral reasoning. The paper by Di Tella et al. (2007) is a truly original study of the effect of exogenous economic events on beliefs about the 'fairness' of the market.

4.1 Gneezy: Consequences

Will lies be told whenever it is beneficial for the liar, regardless of the consequences for the other party?

- Standard economic assumption: Agents only tell the truth when it is in their self-interest to do so. Truth has no intrinsic value. But even economists tell the truth from time to time, without any strategic justification for doing so.
- Questions of this study: Do agents place any intrinsic value on telling the truth? Or, even if not, do they care about the consequences of lying for the other party?
- Four categories of lies (other classifications exist):
  - Lies that help both sides or at least harm no one: "You look great today."
  - Lies that help the other party, even if they harm the liar: Altruism or enjoying the act of giving (but then is it harm?)
  - Lies that do not help the liar but can harm both sides: Spite
  - *Instrumentalism:* Lies that increase the payoff to liar at cost to the other party. This is where all Economic predictions reside.

Idea of this study: Manipulate the benefits of lying for the potential liar, costs to the "lyee" and see what happens.

4.1.1 Method: Cheap-talk sender-receiver game

- Two possible monetary distributions: A or B
- Only player one informed about monetary consequences of each option
- Rules of game known to both participants
- Player one must send one of two messages to player two

    1. Message A: "Option A will earn you more money than option B."
    2. Message B: "Option B will earn you more money than option A."

4.1.2 Predictions

- What does receiver believe?
- What will sender do given these beliefs?
- Is there a Nash equilibrium?
- 82 percent of senders said they expected receivers to follow their message.
- In point of fact, 78 percent of receivers did follow the message.
- Assume senders expect receivers to be credulous (believe what they are told). What will the sender do?
- An additional manipulation: "Previous experience has shown that this recipient always does what s/he is told. Do you want to change your message?" Only 3 of 50 changed message.

4.1.3 Results

- See payoffs in Table 1. See Figure 1. Subjects lie more when the payoff is higher.
- Results so far consistent with either:

1. Dislike lying
2. Care about others' outcomes
3. Both

- Can we tease these apart at all? Re-run the game, but now as a dictator game rather than a game of deception.
- Set exogenous odds of compliance with executing Player 1's choice at 80% (20% chance that other choice taken). (These numbers chosen to correspond to what happened in first game.)
- There is no lying in this treatment. If agents don't mind lying, results should be identical to above. If do mind lying, should appear more self-interested in this treatment.
- With same payoffs, agents are more likely to choose income-maximizing behavior when they do not have to lie.

4.1.4 Conclusions

Gneezy's summary: "The implications of these results are illustrated by the purchase of a car: you can trust what the seller says about the condition of the brakes more than what she says about the state of the air conditioning."

But it's more fundamental in my view: a rejection of 'consequentialism.' Agents care about process as well as payoffs. Do *not* simply value gains to self and losses to other party (cost-benefit). They appear to put weight on how those outcomes arrived at—holding outcomes fixed, like them less if arrived at through deception.

This is a laboratory result, and so due skepticism is warranted. Nevertheless, I suspect there is a robust insight here.

4.2 DiTella, Galiani and Schargrodsky (2007)

This is a one of a kind paper. The context is a historical experiment. In 1981, about 1,800 families occupied ('squatted on') a wasteland area outside of Buenos Aires, Argentina. These

were landless citizens organized by a Catholic priest. To avoid creating a shantytown, they partitioned the land into small parcels. At the time of the occupation, the squatters believed the land belonged to the state. But it turned out to be made up of thirteen tracts of land that belonged to different private owners. After resisting several attempts at removal, the government passed a law in 1984 expropriating the land from its owners (with compensation) and transferring the parcels with legal title to squatters. The owners of 8 tracts accepted the government's compensation offer, and in 1989, settlers on those tracts were given formal titles. The other 5 sued for higher compensation (they couldn't challenge the land transfer, only the terms). In 1998, one of these 5 was settled. The other 4 were ongoing in 2007. Thus, lucky squatters ended up with a title and unlucky squatters did not.

The premise of the paper (and one that appears well documented) is that other than the fortune or misfortune of squatting on the lucky tracts, the different settlers and the quality of the land on which they settled are comparable. The second award (in 1998) provides another treatment/control contrast that is useful in this regard.

Table I summarizes the Intention to Treat setup. Their analysis is based on a survey conducted in 2003 of 448 randomly selected parcels. In 313 households, the first family member had arrived before the lawsuits began. Thus, this group's decision to locate was not endogenous to the allocation outcome (though the decision of others to leave certainly could have been).

4.2.1 Survey questions

1. "Do you believe that it is possible to be successful on your own or a large group that supports each other is necessary?" The two possible answers were "It is possible to be successful on your own" and "A large group is necessary to be successful."

2. "Do you believe that having money is important to be happy?" The possible answers were "Indispensable to be happy," "Very important to be happy," "Important to be happy," and "Not important to be happy."

3. "In general, people who put effort working end up much better, better, worst, or much worst than those that do not put an effort?" The possible answers were "Much better than those that do not put an effort," "Better than those that do not put an effort,"

"Worst than those that do not put an effort," and "Much worst than those that do not put an effort."

4. "In general, in our country, would you say that one can trust other people or that people cannot be trusted?" The possible answers were "You can trust others" and "You cannot trust others."

4.2.2 Results

Table IV shows that land-titling increases the propensity to subscribe to market beliefs in 3 of 4 cases. The effects are quite large. Typically the gap between titled and non-titled settlers is about 20 to 30 percent.

Attrition is a concern, but Table V provides some reassurance. The effects are equally large among those receiving titles in 1998, 17 years after squatting.

Controlling for education, wealth, and access to modern technology (all of which could be the result of property entitlement) does not reduce these relationships. This is not a perfect test of the causal channels, but it is helpful.

Table VIII shows that the effect sizes are large enough to move entitled squatters to the mean market beliefs of the general Buenos Aires population.

4.2.3 Interpretation

Experience hypothesis: Beliefs are formed as a by-product or consequence of agent's activities and interactions with others in society.

Motivated beliefs hypothesis. People change their beliefs to what is convenient to justify their activities.

It's probably not possible to convincingly tease apart these alternatives given the available data in this study.

It's also possible that the negative beliefs of the non-entitled squatters is a result of their bad fortune rather than the market beliefs of entitled squatters being a result of their good fortune (that is, non-entitled squatters were radicalized by their bad experience). This is in some sense a reference point argument. However, the beliefs of the non-entitled appear comparable to

other similarly educated and impoverished squatters elsewhere in Buenos Aires who have not had the bad luck of being denied a title they had expected to receive. Thus, it does not appear that we are simply measuring a scarring effect.

This paper provides a starting point for much more analysis of belief formation. What is most needed is direct evidence on the degree to which beliefs change behavior. At the moment, this is wide open. And the significance of this study ultimately turns on whether or not beliefs 'matter.'

5 Conclusion

This branch of economics regarding norms and fairness has seen very few applications in labor and development economics, outside of canonical tests of efficiency wage theory. But there are real opportunities, as the Mas and Di Tella et al. studies show. The 'fruitful' work by Bandiera et al. on fruit pickers also goes some ways in this direction.

Ernst Fehr asks often, "Why do labor economists still write down market-clearing models?" Do you agree that these models no longer apply?