ELSEVIER

European Economic Review 44 (2000) 1069–1078

EUROPEAN ECONOMIC REVIEW

www.elsevier.com/locate/econbase

Reciprocity and wage undercutting

Martin Dufwenberg[a,*], Georg Kirchsteiger[b]

[a]*Department of Economics, Stockholm University, SE-10691 Stockholm, Sweden*
[b]*Department of Economics, University of Vienna, Vienna, Austria*

Abstract

It is well documented that employers refuse to hire workers who offer their services at less than the prevailing wage. The received explanation is that workers are motivated by reciprocity – they desire to reward kindness and punish hostility. To refuse an outsider's underbid is viewed as a kind choice that is met with good effort; a low wage is viewed as an insult that is met with shirking. We have developed a general theory of reciprocity which in this paper is applied to a wage-setting game played by an employer and two workers. We show that when workers are motivated by reciprocity, equilibrium behaviour accords well with the aforementioned stylized facts.

JEL classification: D63; E24; J41

Keywords: Reciprocity; Wage underbidding; Unemployment

Introduction

A job well done takes a motivated worker. Employers realize this, and may attempt to influence the working morale of their employees. One important instrument in this connection is *the wage*. It may be a good idea to pay a high wage if this makes an employee grateful and prone to work in ways beneficial to the employer. A lower wage, even if it does not make the employee quit his job,

* Corresponding author. Tel.: + 46-8-163547; fax: + 46-8-159482.
E-mail addresses: md@ne.su.se (M. Dufwenberg), georg.kirchsteiger@univie.ac.at (G. Kirchsteiger)

0014-2921/00/$ - see front matter
PII: S0014-2921(99)00047-1

may be regarded as an insult which is met with less conscientious vocational effort. Even in tight labour markets, when unemployment is high, employers may be reluctant to reduce wages for this reason.

This picture is confirmed by scholarly work in many fields. It is accounted for in interview studies that economists have conducted with business leaders (e.g. Agell and Lundborg, 1995, 1999; Bewley, 1998; Blinder and Choi, 1990; Campbell and Kamlani, 1997; Kaufman, 1984). It is supported by experimental labour market studies (e.g. Fehr and Falk, 1998; Fehr et al., 1993, 1998). It is in line with discussions in organization theory (e.g. Steers and Porter, 1991) and psychology (e.g. Argyle, 1989). This work suggests that an important driving force behind the results concerns *reciprocal motivation* – people desire to be kind to anyone they conceive of as kind and to hurt anyone who is unkind. In the case at hand, a worker who receives a high wage thinks of his employer as kind, and the worker is kind in return by exerting lots of effort. Employers avoid hiring people at low wages, foreseeing that this would be conceived of as unkind behaviour that is met with shirking.

In important contributions Akerlof (1982) and Akerlof and Yellen (1990) investigate the economic consequences of such behaviour. However, in their work a positive wage–effort relationship is postulated, so one may wonder if such behaviour will actually emerge endogenously in a model which takes reciprocal motivation as its basic premiss. In a recent paper, Rabin (1993) develops techniques for incorporating reciprocity into game theory and economics. His model is meant to highlight and illustrate qualitative features that are unique to reciprocity though. The model abstracts from information about the sequential structure of a strategic situation, and is therefore not suitable for application to situations with interesting dynamic structures. In a game where decisions about wage offers, hirings and working efforts are taken in turn, the model would not yield sensible predictions.

In Dufwenberg and Kirchsteiger (1998) we develop a theory of reciprocity which is designed for the analysis of the impact of reciprocity on economic problems.[1] The theory is directly inspired by Rabin's work, but works for extensive games in which the sequential structure of a strategic situation is made explicit. It captures the intuitive meaning of reciprocity in situations with a non-trivial dynamic structure, as well as many qualitative features of experimental evidence. In this paper, we apply our model to two wage-setting games played by an employer and two workers. We show that when the workers are motivated by reciprocity, in equilibrium the players' behaviour is consistent with the aforementioned results.

[1] Confer also Falk and Fischbacher (1998) who propose a different approach to modelling reciprocity.

Inspired by experimental results, there also exist approaches designed to investigate not reciprocity, but distributional concerns. These models permit decision makers not only to be motivated by their own payoff, but rather by the final distribution of payoffs. A particular class of these models that have been applied to wage setting games incorporate a desire for a fair allocation, i.e. a person's utility is decreasing in the difference between the own payoff and that of the partner (see e.g. Bolton and Ockenfels, 1999; Fehr et al., 1998; Fehr and Schmidt, 1999). While these fairness approaches are capable of explaining many experimental results,[2] their application to the problem of wage undercutting seems to be more problematic. In most experiments all plausible fairness standards demand the same, namely an equal split allocation (although, of course, subjects do not always behave accordingly). In actual labour relationships, however, it is not clear how to compare the payoff of a firm with the payoff of its workers, and which standards of distributive justice to apply. Should the wages be compared to the profit? If yes, what is a 'fair' relation between wages and profits? If no, what else should be compared? Should shareholders' payoffs arising from an increase in stock-prices be taken into account? Is the gross or the net wage relevant for the comparison? On top of these unsolved questions the information necessary to make 'fairness' evaluations is not available in many cases. Typically, profits of firms as shown in the balance-sheet are shaped by tax avoidance and stock-price considerations. Hence, they often do not reflect the 'true' profits of a firm, and accordingly workers have no good information about it. Similarly, workers are often not informed about labour taxes imposed on the firms. Consequently, workers very often do not even know what their firms have to pay for their labour, i.e. they do not know their actual gross wage. All these informational problems as well as the ambiguities about the relevant fairness concept makes the use of models of distributive justice problematic for the analysis of labour relations.

On the other hand, firms and workers normally know very well the range of possible wages. Hence, they can easily assess the firm's kindness when paying a specific wage. Similarly, the range of possible working efforts, and the kindness of a specific effort level, can be easily evaluated. Hence, contrary to fairness norms, the reciprocity principle – be kind to those who are kind to you – can be easily applied to the analysis of wage undercutting.

2. Results

In this section we consider two wage-setting games and analyse what happens when the workers are motivated by reciprocity. First, however, we give a brief introduction to how the theory of reciprocity in Dufwenberg and Kirchsteiger

[2] See, however, Blount (1995), Charness (1996), or Gneezy et al. (1998) for experimental results that cannot be explained by distributional concerns.

(1998) works. Our ambition is merely to supply some intuition about central ideas. The full theory is somewhat involved, and due to space constraint we must refer to our other paper for details.

Each player i is assumed to choose a strategy that maximizes his utility u_i defined as

$$u_i = \pi_i + Y_i \cdot \sum_{j \neq i} (\kappa_{ij} \cdot \lambda_{iji}).$$

Here π_i is player i's 'material payoff' which represents some objectively measurable quantity, for example money. The term $Y_i \cdot \sum_{j \neq i}(\kappa_{ij} \cdot \lambda_{iji})$ expresses player i's 'reciprocity payoff' with respect to each player $j \neq i$. Y_i is a non-negative parameter describing i's sensitivity to reciprocity. The higher is Y_i, the more sensitive to reciprocity is i. For each $j \neq i$, κ_{ij} represents i's kindness to j. This factor is positive if i is kind, and negative if i is unkind. The factor λ_{iji} represents i's belief about how kind j is to i. It is positive if i believes j is kind to i, and negative if i believes j is unkind to i. The specification captures reciprocity by making it in i's interest to make the sign of κ_{ij} match the sign of λ_{iji}. If $\lambda_{iji} < 0$ player i believes j is unkind to i, and other things being equal i will want to be unkind in return so that $\kappa_{ij} < 0$. Similarly, when $\lambda_{iji} > 0$ player i wants to be kind in return so that $\kappa_{ij} > 0$. This sign matching feature is a key feature of the model.

Another key feature is that κ_{ij} and λ_{iji} depend on player i's *beliefs*. κ_{ij} is measured by comparing the material payoff that i believes that j gets to the set of material payoffs that i believes that j would get were i to choose differently than he does. Effectively, i is kind if he believes he gives j 'a lot' relative to what i believes he could give to j in principle.[3] Conversely, i is unkind if he believes he gives j 'very little' relative to what i believes he could give to j in principle. λ_{iji} is measured similarly, except one has to 'move up' a level in the belief hierarchy: For example, i believes j is kind if i believes j believes she gives i 'a lot' relative to what j believes she could give to i in principle.[4]

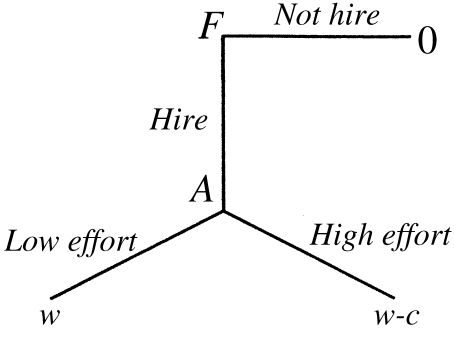
The specification of u_i entails that when i optimizes he may have to make tradeoffs between his reciprocity payoff and his material payoff. In Dufwenberg and Kirchsteiger (1998) we define and prove existence of what we call a *sequential reciprocity equilibrium* (henceforth *SRE*). This concept requires each player i to maximize u_i given correct beliefs. The concept invokes a subgame perfection requirement: all players must optimize in all subgames given strategies and

[3] More precisely, i's kindness is calculated as the difference between what i believes he gives to j, and the average of the maximum and minimum payoff that i believes he could give to j in principle.

[4] Note that, due to κ_{ij} and λ_{iji}, u_i will depend on i's beliefs, unlike in standard games where payoffs depend only on chosen strategies. A general framework for incorporating payoff functions of this form into strategic analysis is *psychological game theory*, introduced by Geanakoplos et al. (1989). Our approach fits into this framework.

beliefs of all players that are updated conditional on the particular subgame under consideration being reached.

Before we turn to the two wage-setting games, it may be helpful to consider a very stylized example with the sole purpose of illustrating the ideas discussed so far. Imagine that a firm F can choose to hire an applicant A at the wage $w > 0$, or not to hire A. If A is hired he subsequently chooses to work with *High effort* or *Low effort*. In the former case A incurs an effort cost of $c > w$. The situation is illustrated in the game G_0. Only A's payoff is specified, given as wage minus cost of effort.



The game G_0

Is F kind if it chooses *Hire*? This depends on F's beliefs. Suppose that F believes that A will choose *Low effort* with probability one. By choosing *Hire* F believes that it gives a payoff of w to A, which can be compared to the payoff of 0 A would get if F chooses *Not hire*. Since $w > 0$, F is kind if it chooses *Hire*, so $\kappa_{FA} > 0$.[5] However, by an analogous argument, one must conclude that F is unkind if it chooses *Hire* while believing that A will choose *High effort*. In this case $\kappa_{FA} < 0$,[6] as F believes it reduces A's payoff as much as possible (from 0 to $w - c < 0$).

In equilibrium, A understands F's motivation. Thus, A's belief about how kind F is to A is given by $\lambda_{AFA} = \kappa_{FA}$. If A cares for reciprocity, how he wants to react depends on the sign of λ_{AFA}. Of course, in equilibrium also F understands A's motivation, and to calculate an SRE one must perform an appropriate fixed point calculation. It could be that no pure strategy profile qualifies, in which case an SRE will involve mixed strategies. To work this out in detail in the case of G_0 we would have to make assumptions about the players' sensitivity to reciprocity as well as the structure of F's material payoffs. However, we leave the

[5] (With reference to footnote 3:) More precisely, in this case $\kappa_{FA} = w - \frac{1}{2}(w + 0) = \frac{1}{2}w$.

[6] (With reference to footnote 3:) More precisely, in this case $\kappa_{FA} = (w - c) - \frac{1}{2}((w - c) + 0) = \frac{1}{2}(w - c)$.

derivation of full solutions for the two economically more interesting wage-setting games which follow now.

Game 1 (*Wage competition*): Imagine a situation where two workers compete to get a job available in a firm. The firm decides whom to hire, and the hired worker then decides about how hard to work. Such a situation can be modelled as a three stage game:

Stage 1: Two applicants simultaneously make wage demands. For simplicity, we assume that a wage demand w can only take two values: $w \in \{w_L, w_H\}$ with $w_L < w_H$.

Stage 2: The firm F accepts one of the demands, denoted by w_A. By that it hires applicant A.

Stage 3: A chooses his work effort e_A, which influences the value of his employment to the firm. For simplicity reasons, we assume that e_A can only take two values: $e_A \in \{e_L, e_H\}$, $e_L < e_H$.

Effort is measured by its impact on F's profit. Thus, F's profit is given by $\pi_F(w_A, e_A) = e_A - w_A$. The cost of low and high effort, respectively, is given by c_L and c_H, $c_L < c_H$, and has to be born by A. Disregarding reciprocity motivation, A's material payoff is $\pi_A(w_A, e_A) = w_A - c_A$. To make the problem interesting, we assume that $e_H - e_L > c_H - c_L$, i.e., the net surplus increases in the effort level. Otherwise, it can never be optimal to choose the high effort level. The material payoff of the rejected applicant is normalized to zero, and we assume the outside option is equally good as getting a low wage and exerting low effort ($\pi_A(w_L, e_L) = w_L - c_L = 0$). (This holds if it is always possible for a rejected applicant to find a low wage-low effort job somewhere else.) Hence, receiving a low wage for a high effort is worse than the outside option ($0 > \pi_A(w_L, e_H) = w_L - c_H$).

To allow for Pareto improvements, we assume wage levels are such that A as well as F gain in terms of material payoff if the high wage is paid for high effort instead of the low wage for low effort ($\pi_A(w_H, e_H) > \pi_A(w_L, e_L)$, $\pi_F(w_H, e_H) > \pi_F(w_L, e_L)$). Hence, a low wage – low effort combination is neither in the interest of A nor F. Yet, in the standard subgame perfect equilibrium where reciprocity plays no role a low wage – low effort combination results. A chooses the low effort level, irrespectively of the wage he receives. Hence, F accepts a low wage demand if feasible.

If the applicants are motivated by reciprocity the outcome is different:[7]

[7] In what follows, the results are driven by the applicants reciprocity motivation towards the firm. If also firms were reciprocally motivated the equilibria we describe would still be valid (and also the firm would experience a reciprocity payoff). Furthermore, the analysis is not affected by an applicant's reciprocity feelings towards the other applicant. For expositional ease (and perhaps also because it is realistic) we proceed the analysis assuming a standard profit maximising firm and no reciprocity concerns between the applicants. Furthermore, we look at the case where both applicants are equally motivated by reciprocity, so that $Y_i = Y$ for any worker i.

Result 1: In every SRE it holds that:

(a) If the firm accepts a low wage demand, the hired worker chooses the low effort level.

(b) If the firm accepts a high wage demand, the hired worker chooses high effort, provided that he is sufficiently motivated by reciprocity, i.e. if $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$.

The intuitive reason for this result is simple:[8] Suppose, contrary to Result 1a, that a low wage demand is accepted and that A responds with a high effort. In equilibrium beliefs are correct, so F must expect a high effort by A. This, however, means that F treats A unkindly, since F believes A's payoff will be lower than zero – the payoff from remaining unemployed. Hence $\kappa_{FA} < 0$. (Do note the central role played by F's belief in justifying the conclusion that F is unkind!) Since in equilibrium beliefs are correct, A understands this, so $\lambda_{AFA} < 0$. With an eye to the definition of u_i above, one sees that A wants to make $\kappa_{AF} < 0$. That is, A wants to be unkind to F in return, and so chooses the low effort level. This is a contradiction. Analogous reasoning shows that if a high wage demand is accepted by F then A is treated kindly even if F expects a high effort. If A is sufficiently inclined to reciprocity he reacts with a high effort choice. Note that the inclination to reciprocity required to get this result ($Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$) is increasing in c_H, decreasing in c_L, decreasing in w_H, and decreasing in the marginal effect of the effort increase $(e_H - e_L)$.

We now restrict our attention to the interesting case where a high effort is enforcable (i.e., $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$), in which case F's equilibrium choice is given by

Result 2: If $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$, in every SRE the firm accepts a high wage demand whenever this is available.

Since a high wage worker provides high effort, the firm's profit is higher if it accepts a high wage demand than a low one. A low wage destroys 'working morale', so the firm does not accept it – wage undercutting does not improve employment prospects.

Game 2 (*Insider vs. outsider*): We now consider a different situation. Imagine that one worker, the insider, is already employed at the high wage w_H, and that an outsider wants to get the insider's job. Such a situation can again be modelled by a three-stage game:

[8] We do not present any formal proofs, but such are available from the authors upon request. However, for interested readers we provide here some information about the derivation of the value of Y, as given in Result 1b: A's payoff from choosing e_H is $\pi_A + Y \cdot \kappa_{AF} \cdot \lambda_{AFA} = (w_H - c_H) + Y \cdot \frac{1}{2}(e_H - e_L) \cdot \frac{1}{2}(w_H - c_H)$; A's payoff from choosing e_L is $(w_H - c_L) + Y \cdot \frac{1}{2}(e_L - e_H) \cdot \frac{1}{2}(w_H - c_H)$. Result 1b holds if the former payoff is larger than the latter, which is equivalent to $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$.

Stage 1: The outsider demands a wage w_O, which can be high or low ($w_O \in \{w_L, w_H\}$).

Stage 2: The firm F accepts or rejects the demand. If it accepts, the outsider is hired at the wage w_O. The (former) insider is then fired and receives the value of the outside option, assumed to be zero. If the firm rejects the outsider's demand, the insider remains employed at the wage w_H. The outsider then remains unemployed, and receives a payoff of zero.

Stage 3: The employed worker, denoted again by A, chooses high or low effort $e_A \in \{e_L, e_H\}$.

We make the same assumptions about effort costs and material payoffs as before, with one addition: If the firm hires the outsider, it has to bear a strictly positive, but arbitrarily small hiring costs T, $0 < T < w_H - w_L$. As one can see below, T serves only as a tie breaking device. (If $T = 0$ additional equilibria can result. On the other hand, if $T > w_H - w_L$, it would never pay to hire the outsider anyhow.)

It is easy to see that again the subgame perfect equilbrium without reciprocity leads to an inefficent low wage-low effort combination. However, if the insider and the outsider are motivated by reciprocity, the outcome is different:

Result 3: In every SRE it holds that:

(a) If the firm accepts a low wage demand, the hired outsider chooses the low effort level.

(b) If the firm accepts a high wage demand, the hired outsider chooses the high effort level, provided that he is sufficiently motivated by reciprocity, i.e if $Y > 2(c_H - c_L)/[(w_H - c_H)(e_h - e_l)]$.

(c) If the firm rejects the outsider's demand, the employed insider chooses high effort, provided that he is sufficiently motivated by reciprocity, i.e if $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$.

Result 4: If $Y > 2(c_H - c_L)/[(w_H - c_H)(e_H - e_L)]$, in every SRE the firm does not hire the outsider.

Due to the reciprocity the insider provides high effort, whereas the outsider provides high effort for the high wage and low effort for the low wage (see Result 3). Hiring the outsider at the high wage is then sub-optimal for the firm, given the hiring cost T.[9] On the other hand, accepting the low wage demand is also not optimal since this would lead to low effort. Hence, wage undercutting does not improve an outsider's employment prospects.

Results 3 and 4 rest on the assumption that the insider's wage is not negotiable. If the insider's wage is flexible, we are back to the framework of

[9] If $T = 0$, hiring the outsider at the high wage as well as sticking to the insider would be part of a SRE. Our main conclusion (wage undercutting does not get the outsider employed) is also with $T = 0$.

Results 1 and 2, where – as we have already seen – wage undercutting is not a promising strategy to get a job. Hence, our main conclusion remains valid irrespectively of whether the wages of the already employed insiders are downward rigid (e.g. by agreements with trade unions) or flexible.

3. Conclusion

Empirical as well as experimental evidence indicates that firms are reluctant to accept low wage offers of workers even in tight labour markets. The employers fear that low paid personnel will not work properly, while a high wage induces a high working morale. We have seen that incorporating reciprocity into two wage-setting games induces behaviour in line with these stylized labour market facts.

Acknowledgements

We thank Mahmood Arai, Sten Nyberg, Klaus Schmidt, a referee, and seminar participants at the 1999 *EEA* meeting in Santiago de Compostela for very helpful comments.

References

Agell, J., Lundborg, P., 1995. Theories of pay and unemployment: Survey evidence from Swedish manufacturing firms. Scandinavian Journal of Economics 97, 295–307.

Agell, J., Lundborg, P., 1999. Survey evidence on wage rigidity: Sweden in the 1990s. Mimeo., Office of Labour Market Evaluations, Uppsala, Sweden.

Akerlof, G.A., 1982. Labour contracts as a partial gift exchange. Quarterly Journal of Economics 97, 543–569.

Akerlof, G.A., Yellen, J.L., 1990. The fair-wage effort hypothesis and unemployment. Quarterly Journal of Economics 195, 255–284.

Argyle, M., 1989. The social psychology of work. 2nd ed. Penguin, Harmondsworth.

Bewley, T., 1998. Why not cut pay? European Economic Review 42, 250–254.

Blinder, A., Choi, D., 1990. A shred of evidence on theories of wage stickiness. Quarterly Journal of Economics 105, 1003–1015.

Blount, S., 1995. When social outcomes aren't fair: The effect of causal attributions on preferences. Organizational Behavior & Human Decision Processes 63 (2), 131–144.

Bolton, G.E., Ockenfels, A., 1999. ERC – A theory of equity, reciprocity and competition, American Economic Review, forthcoming.

Campbell, C.M., Kamlani, K.S., 1997. The reasons for wage rigidity: Evidence from a survey of firms. Quarterly Journal of Economics 112, 759–789.

Charness, G., 1996. Attribution and reciprocity in a simulated labor market: An experimental investigation. Mimeo.

Dufwenberg, M., Kirchsteiger, G., 1998. A theory of sequential reciprocity, CentER Discussion paper No. 9837, Tilburg University.

Falk, A., Fischbacher, U., 1998. A theory of reciprocity. Mimeo., University of Zürich.
Fehr, E., Falk, A., 1998. Wage rigidities in a competitive, incomplete contract market. Journal of Political Economy 107, 106–134.
Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Quarterly Journal of Economics 108 (2), 437–460.
Fehr, E., Kirchsteiger, G., Riedl, A., 1998. Gift exchange and reciprocity in competitive experimental markets. European Economic Review 42 (1), 1–34.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114, 817–868.
Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games and Economic Behavior 1, 60–79.
Gneezy, U., Güth, W., Verboven, F., 1998. Presents or investments? An experimental analysis, CentER Discussion paper No. 9844, Tilburg University.
Kaufman, R., 1984. On wage stickiness in Britain's competitive sector. British Journal of Industrial Relations 22, 101–112.
Rabin, M., 1993. Incooperating fairness into game theory and economics. American Economic Review 83, 1281–1302.
Steers, R.M., Porter, L.W., 1991. Motivation and work behavior, 5th ed. McGraw-Hill, New York.