Internal incentive mechanisms in small teams

with union-firm wage bargaining

By

Jon Strand

Department of Economics

University of Oslo

Box 1095, Blindern, 0317 Oslo, Norway

Jon.strand@econ.uio.no

November 25, 2004

Very preliminary: Do not quote or cite

Abstract

I study firms that consist of small teams where individual workers' contributions to output are observable only within each team, team members interact continuously over time and each team bargains with the firm over the net value of output. I assume that trigger strategies are used within teams to support more efficient outcomes than the non-cooperative one. The wage bargaining solution is shown to imply a wage share between ½ and unity of output, depending on relative bargaining powers, and is more efficient (gives better worker effort incentives) when the wage share is higher as effort incentives are better. The best achievable solution is shown to be less efficient when team size is larger, teams have shorter duration, and with greater turnover of team leader and other team personnel. This implies tradeoffs between team turnover and scale economies from larger team size on one hand, and team member incentives on the other.

Key words: Team production; moral hazard; union-firm wage bargaining; effort enforcement.

JEL classification: D82, J31, J53

1. Introduction

This paper studies intra-firm incentives under team production. A *team* I defined by Holmström (1982) as "a group of individuals who are organized so that their productive inputs are related". Team production is ubiquitous in modern economies, and involves economies of scale whereby output per worker increases over some range for team size.[1] A disadvantage of team relative to individualized production is that observation and identification of each member's contribution to output is usually more costly and often infeasible. As a consequence individual worker reward must depend to a large extent on measures of (observable) aggregate group output instead of individual worker output. As emphasized by Holmström and later several others, this often creates moral hazard whereby workers have incentives to put up too little effort. Team work may however also imply benefits, in terms of better cooperation and sense of responsibility among workers, improvements in work environment (Ferreira 2002), and improved information transfer between co-workers (Lazear 1998).

The stated moral hazard problem can be countered in several ways. One way is strict performance monitoring of teams and team members by the firm. As I have argued elsewhere (Strand 2004) this may however involve substantial monitoring costs, and in addition other costs as individual shirking workers may put up additional "effort" to avoid being discovered. Moreover, when workers are unionized as will be assumed here, firms' real authority to monitor may be limited by existing union-firm contracts; and firms may have little room for individual differentiation of wages between hard-working and shirking workers, or for firing alledged shirkers.

[1] See e.g. Marino and Zabojnik (2001) for further discussions and analysis of scale economies in teams.

A more relevant alternative may then be that monitoring takes place among workers themselves A basic assumption in this paper is that there are benefits to cooperation between workers within teams, and that the costs of observing and monitoring individual worker efforts lower for workers themselves than for the firm's owners (or owners' agents such as managers and supervisors), than for workers themselves. Such a monitoring role to workers may seem to be particularly important when, as I assume here, workers are unionized and bargain with the firm over the split of overall output produced by each team, of given size n, where team output is observable. We then invoke the so-called "right-to-manage" model where the firm is formally in charge of the employment decision, while there is bargaining over the split of net output. In such a union-firm setting the firm is likely to able to use strong group sanctions of the type prescribed in Holmström (1982), but is instead confined to using simpler (and more practically-implementable) partnership sharing rules similar to those of Kandel and Lazear (1992). I assume that output arises simply as an aggregate of group members' output (where individual workers' outputs are formally independent within the team), and that purely moral sanctions or motivations play no role in determining workers' effort choices. A lack of cooperation within the group then leads to moral hazard in reducing output substantially below the efficient level. Effort incentives (and output) are affected both by the sharing rule between firms and the union, and by the incentive mechanism used among workers themselves. I assume this mechanism to be of the trigger strategy type whereby a given (equilibrium) degree of cooperation breaks down if one worker is identified as shirking (or not abide by the cooperative agreement) by at least one co-worker. This strategy is shown to always support a greater degree of cooperation than the purely non-cooperative outcome, but is often not fully efficient. I also derive an equilibrium sharing rule between the firm and the workers as a group. The share β in net output going to workers is shown to always lie between ½ and unity, depending on the bargaining powers of the two parties, and equals ¾ in the symmetric Nash case (equal bargaining powers). β is independent of the achieved degree of cooperation within the team. The firm's unconstrained optimal share (1-β) here equals ½, which optimally balances the profit share and overall output via workers' effort incentives.

The paper is related to a sizeable literature on internal team production and incentives. Our notion of intra-team enforcement is closely related to the concept of "peer pressure" as formalized by Kandel and Lazear (1992) and Barron and Gjerde (1997), where co-workers can exert non-pecuniary pressure on a (potentially) shirking worker so as to induce higher worker's effort or removing possible shirking incentives. Backes-Gellner et al (2004) consider team size for start-up firm partnerships, and argue that the incentive effects to overcome moral hazard are typically weakened when the team grows larger (beyond very small sizes); in their theoretical example the optimal team size is 3. Potentially, better effort enforcement may be possible when there are more co-workers who all may control each other against shirking. Backes-Gellner et al however argue that the increased moral hazard effect on effort when team size grows from a small number will typically dominate potential benefits. I here reach similar conclusions for team size effects, from a model with a more traditional economic enforcement mechanism (intra-group collusion through repeated interaction between agents): I find this type of mechanism even more relevant in the context of a worker group within a firm, than in the entrepreneurial context of start-up firms.

In spirit this paper is close to Che and Yoo (2001) where a principal designs an optimal incentive scheme within a two-agent team. They show that repeated interaction and “high” degrees of mutual observation between the two agents in a repeated setting gives rise to highly efficient solutions supported by trigger strategies. The main difference from this paper is in the imposed institutional structure of union-management relationships governed by the “right-to-manage” solution whereby firms have direct control over employment and employment rules but not over wages or wage schemes. Incentives are here at the outset further diluted by the inability of firm management to control the wage incentive mechanisms. Instead, the firm’s management controls other variables (such as worker turnover and team size) that affect the set of implementable solutions, which is not considered in Che and Yoo (2001). The endogenous derivation of the firm-worker bargaining solution is a further added feature of our model.

In the sequel, the model is presented in section 2, first the intra-team incentive mechanism in subsection 2.1, and then the management-union bargaining solution and the overall equilibrium in subsection 2.2. Section 3 concludes and discusses some practical implications for firm personnel policy and efficiency.

2. The model

Consider a team of given size (n>1) where only joint output is observed outside of the team itself. Group members are organized in a (local) union that bargains with the firm over a net surplus from the joint production of the team. This firm-union game structure puts strict limits on allowable types of effort enforcement.[2] In particular, enforcement mechanisms considered by Holmström (1982), whereby the group as a whole is "punished" for not attaining a possible required group output, cannot be used.[3] This leads to a "low-powered" incentive mechanism which is quite different from mechanisms prescribed by traditional principal-agent theory, e.g. Hart and Holmström (1987). It creates a need for alternative enforcement mechanisms within the group. Such a mechanism, focused on here, is *mutual monitoring among team members*, in a manner similar to Kandel and Lazear (1992), Che and Yoo (2001) and Towry (2003), where team members are assumed to have (at least potentially) a long-lasting relationship. Each member perfectly controls ones own contribution to overall group output, and in addition has some imperfect technology for observing *other members'* outputs. This may be used as a basis for *internal group punishment and enforcement*.[4]

The model is developed in two separate steps, corresponding to the two stages of the worker-firm game specified. Subsection 2.1 first deals with the production stage, where group output is determined for a given reward function (as given from the second stage). I here build a dynamic model for internal enforcement of efforts in an n-person partnership. The second is the bargaining (or wage formation) stage, dealt with in

[2] One might perhaps even argue that in this institutional setting, individual outputs can be observable to the firm's management without management having the opportunity to utilize such information for better effort enforcement. Such observation might however create incentives for collusive arrangements between management and the union, beyond those discussed here.

[3] While technically possible under the informational assumptions made by Holmström, his enforcement mechanisms are hardly ever seen in practice and are rather unthinkable in a setting with labor unions, typical at least in a European context. Our point of departure is here closer to that of Kandel and Lazear (1992), where the team is viewed as a sharing partnership which formally rules out Holmström's mechanisms.

[4] See Che and Yoo (2001) for a discussion of similar mechansmis, and several examples of applications of such mechanisms from U.S. industry.

subsection 2.2. I here specify a bargaining model, where workers as a group bargain over the general wage level with the firm.

Note that the standard backward induction procedure for solving similar sequential games, here need not be applied, as the principal solutions under the two steps are independent. In terms of exposition it is advantageous to present the solution in "forward induction" style, where equilibrium effort is determined first, and the wage subsequently.

2.1. Internal partnership incentives

Assume that the firm in question has a number of teams whose outputs are independent. Each team consists of n identical workers, with utility functions

$$u_i = w_i - e(y_i), i = 1, ..., n. \quad (1)$$

where w_i is the wage paid to group member i, y_i the output produced by this group member, and e is subjective effort associated with producing output y_i (consequently all workers have identical preferences with respect to effort). Workers are assumed to be risk neutral and have no value of leisure of unemployment benefits when out of work; thus their short-run opportunity cost of working (reservation wage) equals zero. Each team member is averse to effort, whereby subjective effort is a quadratic function of individual output level y_i:

$$e(y_i) = \frac{1}{2b} y_i^2. \quad (2)$$

Team output is Y = Σy_i.[5] In the absence of uncertainty the contribution of worker i to overall group output can then in principle be identified. Assume that team members collectively receive a given fraction βy of total group output in the form of wages, where β is a parameter characterizing the outcome of workers' collective bargaining power versus the firm. We will later find (in subsection 2.2 below) that β can be taken as exogenous at the effort-determination stage. Team members are identical and share wages equally among themselves. The wage to a given worker i is then

$$w_i = \beta \frac{Y}{n} = \beta \frac{\sum_{i=1}^{n} y_i}{n} \tag{3}$$

Given that workers receive a fixed fraction β of total output, we may define a "social optimum" for workers, from maximizing the function

$$V = \sum y_i - \sum \beta \frac{1}{2b} (y_i)^2 \tag{4}$$

with respect to y_i The first order condition yields

$$\frac{\partial V}{\partial y_i^c} = \beta - \frac{1}{b} y_i(c) = 0 \Rightarrow y_i = y_i(c) = \beta b \tag{5}$$

[5] Thus there are no "synergies" in production. This assumption is made mainly for analytical simplicity, in order to focus on other things than those arising from "synergies" as such. One could here however legitimately ask what is here the point of team production. Our answer to this is to simply postulate that a groups size n is necessary for efficient production (i.e., if group size were reduced below n, output would fall drastically).

The current wage and worker utility levels equal

$$w(c) = \beta^2 b\,. \tag{6}$$

$$u(c) = \frac{\beta^2 b}{2} \tag{6a}$$

This is a "benchmark" solution implemented when the n workers jointly determine effort levels efficiently. It may be contrasted to two other benchmark-type solutions. The first of these is the non-cooperative Nash equilibrium solution, where each worker sets his or her individually optimal level of $y_i = y_i(0)$, taking as given the other workers' levels of y, $y_j(e)$. The worker then maximizes the following function:

$$u_i(0) = \beta \frac{y_i(0) + \sum_{j \neq i} y_j(e)}{n} - \frac{1}{2b}(y_i(0))^2 \tag{7}$$

The first-order condition for worker i implies the following rule for $y_i(0)$:

$$y_i(0) = \frac{\beta b}{n} \tag{8}$$

This choice is made identically by all workers. In equilibrium, $y_j = y_i(0)$ from (8). From (5) and (8), the joint non-cooperative equilibrium group output is here reduced to only 1/n relative to the efficient level for the group as a whole. The corresponding wage and utility levels are then

$$w_i(0) = \frac{\beta^2 b}{n} \tag{9}$$

$$u(0) = \frac{\beta^2 b(2n-1)}{2n^2} \tag{10}$$

For all n > 1, u(0) < u(c).

The other extreme benchmark solution is β=1 in (6) and (6a), which implies that workers are residual claimants and at the same time cooperate fully. This leads to a fully efficient (first-best) solution where y(e) = w(e) = b, and u(e) = b/2. This solution can however safely be ruled out here as long as β actually is smaller than unity, as workers can never be provided with incentives to set e at this level.

A basic problem for workers is to overcome the free-rider incentives associated with imperfect effort enforcement. I here consider an internal enforcement mechanism based on workers policing or controlling each other. In particular, consider a possible symmetric equilibrium where output per worker and wage are given by

$$y(\alpha) = \frac{\beta b}{\alpha} \tag{10a}$$

$$w(\alpha) = \frac{\beta^2 b}{\alpha}, \; 1 \le \alpha \le n. \tag{10b}$$

The fully cooperative solution among workers, (6), may here be “too demanding”, and one may need to settle for a second-best intermediate between the non-cooperative and fully cooperative ones. The current worker utility level is

$$u(\alpha)=\frac{\beta^2 b(2\alpha-1)}{2\alpha^2}, \alpha\in[1,n]. \tag{11}$$

For 1<α<n we have u(0)<u(α)<u(c). This implies gains due to (imperfect) cooperation over no cooperation, although smaller than the potential gains under full cooperation (which corresponds to setting α=1).

The enforcement mechanism we consider is based on workers imperfectly observing each others' efforts and reacting by "punishments" when shirking by any one worker is observed. Firm management plays no direct role in effort enforcement.

To test the stability of the established cooperative solution, consider a deviation by one team member from this solution, in which case worker in question chooses to produce at rate y(d). Assume that any deviation has the same probability of being discovered.[6] We then find that y(d) = y(0). The worker's wage and utility are

$$w(d)=\frac{\beta^2 b}{\alpha n^2}\left[n(n-1)+\alpha\right] \tag{12}$$

$$u(d)=\frac{\beta^2 b}{2\alpha n^2}\left[2n(n-1)+\alpha\right] \tag{13}$$

The current utility gain from deviating optimally from the established cooperative solution is given by

[6] This corresponds to the type of enforcement mechanism used by Shapiro and Stiglitz (1984), but is clearly somewhat unrealistic here. The underlying mechanism here must involve an assumption that individual workers' outputs are not perfectly observable by co-workers, which in turn is likely to imply that a "large" deviation can more easily be discovered than a "small" deviation. I leave considerations for such variable observability to future work.

$$u(d) - u(\alpha) = \frac{(n-\alpha)^2 \beta^2 b}{2\alpha^2 n^2} \tag{14}$$

Consider the following value functions (identical for all workers):

V(α) = the lifetime discounted value of currently (and in the future) producing y(α), given that all other workers retain the same level of output.

V(d) = the lifetime discounted value of defecting from the above solution (currently "shirking" and instead producing y(d), while the other workers are producing y(α));

V(0) = the lifetime discounted value of current non-cooperation (where all workers are "defecting");

V(u) = the lifetime value of currently being unemployed.

Define also the following parameters:

q(n) = the (continuous) rate at which shirking by one worker is detected and punished by other team members (the time until detection is exponentially distributed with expectation 1/q). Punishment takes the form that an initially cooperative solution breaks down and a non-cooperative solution among the n team members takes its place (until again altered as discussed below). q most certainly varies with team size n, although the functional relationship is not obvious. With a larger team, the number of co-workers that can observe an individual deviation is larger, and this might lead to a higher overall probability of detection. Shirking may still have a smaller probability of actually being punished by co-workers when a team is larger, for at least two separate reasons. First, an output deviation by one worker could be less noticeable to any other worker when the team is large. Secondly, in contrast to the two-person team case, a co-worker who discovers shirking may not necessarily wish to sanction the deviation when the team is large (and may then instead e.g. seek to collude in more complicated ways with the other co-worker). We here simply assume that the two last arguments outweigh the first, and that q decreases in n. (Note also that for given n, a smaller α implies a large n-α, i.e., a large relative deviation in output when shirking takes place, which may imply that shirking is more noticeable and consequently easier to punish.)

λ(n) = the rate at which the initially non-cooperative solution is restored following breakdown of cooperation. This could result from sporadic renegotiation between the n members, or some event triggered either by exogenous events or by the firm (such as replacement of a team leader). λ is likely to be a function of n, although the effect cannot obviously be signed.[7]

μ = the rate at which the current (potentially deviating) worker is removed from the current team and placed in another team in the same firm.

ρ = the rate at which the worker moves to another firm (without any intermittent period of unemployment).

σ = the rate at which the worker loses his job in the current firm, due to a team breakup.

θ = the rate at which an unemployed worker finds a new job equivalent to the one described.

b = the current utility of being unemployed.

Among these parameters, q, ρ and θ can be viewed as under control by the worker, n, μ and σ by the firm, while both are likely to have influence on λ.

[7] It is far from obvious whether it becomes more or less difficult to reestablish cooperation, when the team is larger. Sporadic reestablishment may be easier for a smaller team, while the firm on the other hand may feel to have greater need to help in reestablishing cooperation when the team is large, e.g. by changing team leader.

Assuming a constant and common rate of interest r, the following set of value functions can now be constructed:

$$rV(\alpha) = u(\alpha) + \sigma(V(u) - V(\alpha)) \tag{15}$$

$$rV(d) = u(d) + q(n)(V(0) - V(d)) + (\mu + \rho)(V(\alpha) - V(d)) + \sigma(V(u) - V(d)) \tag{16}$$

$$rV(0) = u(0) + \lambda(n)(V(\alpha) - V(0)) + (\mu + \rho)(V(\alpha) - V(0)) + \sigma(V(u) - V(0)) \tag{17}$$

$$rV(u) = b + \theta(V(\alpha) - V(u)) \tag{18}$$

In addition we need to specify the non-shirking constraint V(α) ≥ V(d). (15) and (18) solve for V(α) and V(u) and their difference, as follows:

$$V(\alpha) = \frac{1}{r}\left[\frac{r+\theta}{r+\sigma+\theta}u(\alpha) + \frac{\sigma}{r+\sigma+\theta}b\right] \tag{19}$$

$$V(u) = \frac{1}{r}\left[\frac{\theta}{r+\sigma+\theta}u(\alpha) + \frac{r+\sigma}{r+\sigma+\theta}b\right] \tag{20}$$

$$V(\alpha) - V(u) = \frac{1}{r+\sigma+\theta}(u(\alpha) - b)) \tag{21}$$

where u(α) is given from (11). The non-shirking constraint can be expressed as

$$u(d) - u(\alpha) \le q(n)(V(\alpha) - V(0)) = \frac{q(n)}{r + \lambda + \mu + \rho + \sigma}(u(\alpha) - u(0)) \tag{22}$$

This is equivalent to the following condition on α:

$$\alpha \ge \frac{q(n) + r + \lambda + \mu + \rho + \sigma}{(2n-1)q(n) + r + \lambda + \mu + \rho + \sigma} n \tag{23}$$

Setting A = r+λ+μ+ρ+σ, enforcement of efficient effort as viewed by workers as a group (α = 1) is feasible whenever

$$\frac{q(n)}{A} \ge \frac{n-1}{3n-1} \tag{24}$$

I will call a team is <u>efficient</u> if it implements the efficient level of effort among team members. The results just derived permits the following claim:

<u>Proposition 1:</u> Assume that q(n) is non-increasing in n and that A is constant. Then a team with n ≤ n members is <u>efficient</u>, and a team with n > n* members is not efficient, where n* is the level of n that yields equality in (24).*

Proposition 1 is easily derived from the property that (n-1)/(3n-1) is strictly increasing in n. It implies that a small team is under our assumptions “more likely” to be efficient than a large one. The proposition has real implications only when n* ≥ 2. Note also that the model makes no claims as to what team sizes are technically efficient, as the analysis throughout is conditional on team size.

An interesting aspect of this result is that the parameters r, λ, μ, ρ and σ all enter into the relationship (23), determining both the limiting and actual values for α, in a simple additive way, even though they have very diverse interpretations and arise from different economic rationales and reasons. Thus in this context, for the firm in controlling the parameters the only concern here ought to be for the sum of μ+σ and not their separate values, which may be important in a policy context.

I will now look at bit closer at the cases where (24) does not hold (for relevant n values), and workers select the lowest possible value of α (> 1) consistent with enforcement of the non-shirking constraint. Such solutions will be inefficient. The relationship between α and n is found differentiating (23) with equality:

$$\frac{d\alpha}{dn} = \frac{(A-q)(A+q) - n(n-1)Aq'}{((2n-1)q + A)^2} \tag{25}$$

Assuming as above that q' ≤ 0, and considering only cases where (24) does not hold (implying A>q), α increases in n. The most favorable cooperative solution then becomes less favorable when n grows larger.

This however does not necessarily imply that the <u>workers' gains from cooperating</u> (relative to the corresponding benchmark non-cooperative solution) is reduced with n, since from (10) above, also the non-cooperative solution becomes less efficient when n grows. The utility gain from cooperating instead of not cooperating changes with n according to the following formula:

$$\frac{d}{dn}\Delta u(\alpha,n)=\frac{d}{dn}[u(\alpha)-u(0)]=\beta^2 b\left[\frac{n-1}{n^3}-\frac{\alpha-1}{\alpha^3}\alpha'\right], \tag{26}$$

where α’ denotes dα/dn, from (25). When α is very close to n (and the solution is close to the non-cooperative one), the expression on the right-hand side of (26) is most certainly positive (since α’ is substantially smaller than unity). On the other hand, when α is much smaller than n (and the best implementable solution is more efficient), this expression could easily be negative.

In line with standard assumptions behind the right-to-manage model, the firm can as noted be viewed as controlling μ (the frequency of moving workers between teams), σ (the job loss rate for workers) and n (team size).[8] For given n, from (23) (with equality) α increases in both μ and σ. As already shown, for given other parameters α increases in n. We here do not directly model the efficiency effects of possible changes in n, μ and σ for the firm. But our analysis at least shows that increasing these parameters is costly in the sense that cooperative output among team members drops. An extended model would here be required to analyze these tradeoffs more explicitly.

Another important point to notice about this solution is that best implementable α values are independent of the bargaining parameter β. From (10a), output itself is however proportional to β.

---

[8] From the discussion in the text above, the firm could possibly also be viewed to control λ (the rate at which cooperation between co-workers in a team is restored after it has broken down), by controlling e.g. the rate at which the “team leader” is replaced. We will not emphasize this effect here.

2.2 The bargaining solution

I now proceed to an analysis of the bargaining stage where the firm and workers as a group bargain over joint output as given by (10a), assuming as before that the "right-to-manage" model is in force. The firm and the workers as a group then are assumed to split overall output in shares 1-β and β, respectively.[9] From subsection 2.1 there is no relationship between the efficiency parameter α and the bargaining parameter β. This implies that there is no technical problem with taking α as exogenous in this exercise. Firm profit per worker is given by

$$\pi(\alpha,\beta)=\frac{\beta(1-\beta)b}{\alpha}\ , \tag{27}$$

where α as before is assumed to be given from (23). The split parameter β has up until now been considered as exogenous. The present analysis aims to determine this parameter, as part of a Nash wage bargaining solution where the two parties start out with relative (Nash cooperative) bargaining strengths γ and 1-γ for workers and firm, respectively. I assume that bargaining only is possible over the total ex post output "cake", which in particular implies that including workers' effort in the bargaining solution is not relevant. The problem of determining an optimal level of β is here simplified substantially by our result that the best implementable solution among workers, represented by α, is independent of β.

[9] The assumption of such a split may not seem fully consistent with standard bargaining theory (with continuous renegotiable bargaining), where, presumably, workers' gain from bargaining ought to be taken as net of effort cost. The formulation here is rather consistent with an assumption that worker effort is "sunk" at the time bargaining takes place, or that it cannot be verified.

As a preliminary consideration I will study whether there exist "ideal" values of β for the two parties, and what these values are. For workers this is simple: β(W)=1 is obviously ideal (letting W denote the worker-preferred value), as worker utility from (11) is maximized for this value (both output and workers' share increase strictly in β). For firms, the preferred value of β can be found differentiating (27) with respect to β (and again noting that α is independent of β), giving the solution for β as follows (call it β(F)):

$$\beta(F) = \frac{1}{2} \tag{28}$$

The firm thus prefers an equal split between the two parties. The reason is that the share of total output going to the firm obviously drops in β, while output increases in β, implying that an equal split between workers and firm is optimal for the firm.

Define the Nash product

$$N(\beta;\alpha) = \left[\frac{\beta^2 b(2\alpha-1)}{2\alpha^2}\right]^{\gamma}\left[\frac{\beta(1-\beta)b}{\alpha}\right]^{1-\gamma}, \tag{29}$$

where the arguments are respectively each worker's net utility, and the firm's profit per worker, as functions of β. Each worker here has a short-run opportunity value of working equal to zero, something that defines the worker's impasse point.[10] N(β,α) is now

[10] We invoke the Rubinstein (1982) non-cooperative bargaining model whereby the short-run alternatives of the two parties are defined by the values of perpetual conflict, here equal to zero for both; see also Muthoo (1999) chapter 3.

maximized with respect to β. This leads to the following simple solution for β (denoted β(N)):

$$\beta(N) = \frac{1+\gamma}{2} \tag{30}$$

This leads immediately to the following result:

<u>Proposition 2:</u> *Assume that the conditions for Proposition 1 hold, and that workers and firm engage in Nash bargaining with relative bargaining strengths γ and 1-γ to workers and firm respectively. Then the relative bargaining strengths of workers and firm are (1+γ)/2 and 1-γ)/2 respectively. The per-worker solutions for output y(α;γ), utility u(α;γ) and profit π(α;γ) are given by*

$$y(\alpha;\gamma) = \frac{(1+\gamma)b}{2\alpha} \tag{31}$$

$$u(\alpha;\gamma) = \frac{(1+\gamma)^2(2\alpha-1)b}{8\alpha^2} \tag{32}$$

$$\pi(\alpha;\gamma) = \frac{(1-\gamma^2)b}{4\alpha}. \tag{33}$$

From Proposition 2, we see in particular that when the two parties have equal basic bargaining strengths ($\gamma = ½$), the share $\beta(N)$ in net ex post output going to workers equals ¾, which is of course midway between the respective "ideal" solutions, $\beta(F)$ and $\beta(W)$.

(31) shows that higher bargaining strength to workers raises output, and in addition makes the solution more efficient by raising workers' efforts. The measure of aggregate per-worker utilities of worker and firm is

$$V(\alpha;\gamma) = \frac{4\alpha - 1 - \gamma}{8\alpha^2}(1+\gamma)b \tag{34}$$

which is an increasing function of γ. While firm profits (of course) drop when β increases beyond ½, each worker's net welfare increases by even more, due to both the direct revenue effect on the wage, and to the advantageous incentive effect on effort.

An interesting implication of Propositions 1-2 taken together is that the bargaining share β going to workers is independent of the internal team effort allocation solution, and depends only on the relative bargaining strength parameter γ. This just confirms our previous assertion about the appropriateness of treating the worker effort allocation problem as independent of the bargaining problem. Note however that this is not a very general result. It follows from our particular specification, importantly, that workers have no utility of leisure. A more general specification of the utility functions would lead to interactions between the bargaining outcome and the effort allocation solution, and thus an overall more complex solution.

3. Conclusions and final comments

We have in this paper studied a model of firm production organized in teams each consisting of identical workers in given number n. Individual workers' contributions to output can be observed only by workers themselves, within each team. The workers in each team are organized and bargain with the firm over the value of total team output. The firm has otherwise no influence on the wage or worker incentives except through general personnel policies such as the determination of team size and turnover, over which the firm has full control.

The analytical solution can be considered in two steps, the worker effort allocation step, and the wage bargaining step, which are shown to be independent given that workers have no value of leisure. The effort allocation, in step 1, is based on workers themselves using trigger strategies to abandon a given degree of cooperation whenever shirking by one worker is detected by a co-worker. This supports a solution which is superior to the static Nash equilibrium effort allocation, but always inferior to the efficient allocation, where the degree of inefficiency also depends on the bargaining solution in step 2. A higher and more efficient level of effort is supported when team size is small, the breakup rate of teams is low, and when there is little turnover among team members, in particular for the team leader (whenever there exists one).

In step 2, the firm and team members as a group bargain over the net surplus from production (in continuous time, and with continuous renegotiation), recognizing that the bargaining solution has implications for the effort chosen by team members in step 1. We then show that the equilibrium wage share in net output is always between ½ and 1, and proportional to the relative worker bargaining strength on this interval. The reason for this result is that the "best" wage share for the firm is ½, which provides the optimal tradeoff for the firm between additional immediate profits for given effort, and workers' best implementable effort. We also find that the overall solution becomes more efficient as the share β going to workers increases.

The model and its results have some positive and some normative implications for firm structure and policies, given that firms' labor forces are unionized, the basic structure of decision making within the firm is of the "right-to-manage" type, and firms organize their workers in teams of the types assumed. First, there are implications for team size, although these implications are somewhat limited within the model. The only formalized effect of team size in the model is via the best implementable effort within each team. Here the main conclusion is that a large team will tend to give less efficient effort incentives than a smaller one, at least beyond a certain minimum team size. There are several possible reasons for this, some within and some outside of the model. First, the "benchmark" non-cooperative solution is more efficient with a small team, implying that less collusion is required to attain efficiency in small teams. Secondly, in a small team each worker's effort is more significant, giving each worker more incentives to monitor any one co-worker, and punishing a deviation. In the same vein, incentives to deviate collusively from an established cooperative solution, among a small group of team members, may exist in a large but not a small team. In general, this result points to tradeoffs between firm size and implementable effort within teams. An implication ought to be that teams are (or ought to be) organized at smaller than technically optimal scales.

Firms' turnover policies are also shown to be of importance in such a setting. Turnover is shown to affect the degree of positive collusion over effort that is feasible within the team group, in at least three separate ways. First, the firm may want to reduce the time period that any one team member stays with a given team, for efficiency reasons (in the model, by increasing the parameter μ). This is however reduces the degree of feasible intra-team cooperation by making between-worker relationships shorter lasting. Secondly, a higher degree of break-up of teams, implying that a given team on average is retained for a shorter period of time (represented by an increased σ), has the same type of effect. Thirdly, one may visualize "team leaders" as individuals who are responsible for establishing cooperation within the group (or at least attempting to do so). Thus, when cooperation has broken down due to an enforced punishment path after shirking, the arrival of a new team leader may lead to reestablishment of cooperation. We show that a greater degree of such reestablishment (represented by an increase in the parameter λ) is damaging to effort incentives, by reducing the expected loss associated with a given punishment path. This is formally similar to the idea of renegotiation of a non-cooperative punishment path solution, with similar effects; see e.g. Farrell and Maskin (1989).

A number of economic variables that are normally outside of the control of individual workers and firms also affect the derived solutions. First, a higher rate of outside hiring by other firms (represented in the model by the parameter ρ) leads to higher general turnover that makes cooperation over effort among team members more difficult. Secondly, one might here also naturally visualize effects via changes in the general level of unemployment in the economy, which in turn affects the rate at which unemployed workers locate new jobs, represented in the model by the parameter θ. There are no formal effects on incentives due to changes in θ in the model, which in turn is due to the assumption that there are no effects whatsoever of a worker's shirking on the probability of retaining or losing ones job. This in turn follows from the (extreme) assumption that there is no communication from the team group to the firm's management, concerning what individuals may or may not be shirking. In a right-to-manage setting, with the firm fully in control of personnel policy, there are is then no ground for dismissals (or promotions) by performance, which is of course unrealistic.

Another result is that the overall solution becomes more efficient when workers as a group receive a larger fraction of overall ex post output. This is due to the moral hazard effect of reducing workers' share, and which in turn is due to the rigid structure of rewards that follows from the right-to-manage model, where the union bargains over total wage payments and all workers are to receive the same wage. But this result also ignores other factors, in particular the possibility that firm entry and exit may be affected. (More on this.)

The analysis here also indicates some interesting avenues for future work. These are both theoretical and empirical. On the theoretical side, one may want more clearly and realistically defined roles for the firm and workers in effort enforcement, and not as here, with workers doing the entire job. The issue of monitoring technology them becomes central. Here, the technology by which firms monitor workers is very primitive and implies no such possibilities. In reality, all firms conduct some amount of monitoring of individual workers, but its structure and optimal amount may vary. On the empirical side, the issue of possible tradeoffs, between technical team size efficiency (which may draw in the direction of large teams) and internal team incentives (acting to reduce them), is an interesting topic. The same is the study of what motivates team members to put up effort (beyond some necessary minimum). Much of the recent literature (cf. e.g. Camerer

(2003), Fehr and Schmidt (2003), Fehr, Klein and Schmidt (2004)) here focuses on the relationship between fairness and incentives .One here shows that substantial fractions of individuals typically select efforts close to efficient levels even in the absence of explicit enforcement, given that the chosen wage system is deemed "fair", usually taken to mean that wage differences are small. In our model, of course, the wage system is "fair"; we do however require an effort enforcement mechanism albeit a somewhat untraditional one. Some of the relevant empirical literature, such as Lazear (2000), Paarsch and Shearer (2000) and Shearer (2004), has demonstrated that there typically are positive effects of explicit incentives on effort, although they are often not particularly strong.

The analysis here also has implications for our view of the role of labor unions and union-firm interrelations, given a "right-to-manage" union-firm relationship; see Booth (1995) for a presentation and discussion. Union-firm contracts typically place severe constraints on firms' ability to design incentive mechanisms for overcoming internal moral hazard problems, due both to limited ability both to differentiate wages and to fire workers for cause. This gives a major role for the union in monitoring and enforcing workers' efforts, out of self-interest. While this apperas plausible in principle, there is little empirical evidence on the issue of union- or union-member-initiated monitoring, and how it relates to firm-initiated monitoring. Clearly, more research is needed in all these areas.

References

Backes-Gellner, U., Mohnen, A. and Werner, A. (2004), Team size and effort in start-up teams – another consequence of free-riding and peer pressure in partnerships. Working paper, University of Zurich.

Barron, J. M. and Gjerde, K. P. (1997), Peer pressure in an agency relationship. *Journal of Labor Economics*, 15, 234-254.

Booth, A. L. (1995), *The economics of the trade union*. Cambridge: Cambridge University Press.

Camerer, C. F. (2003), *Behavioral game theory: Experiments in strategic interaction.* Princeton, N. J.: Princeton University Press.

CheY. K. and Yoo, S. W. (2001), Optimal incentives for teams. *American Economic Review*, 91, 525-541.

Farrell, J. and Maskin, E. (1989), *Games and Economic Behavior*, 1, 327-360.

Fehr, E. and Schimdt, K. M. (2003), Theories of fairnessand reciprocity: Evidence and economc applications. In M. Dewatripont et al: *Advances in Economic Theory, Eigth World Congress of the Econometric Society*, vol 1. Cambridge: Cambridge University Press.

Fehr, E. and Schmidt, K. M. (2004), Contracts, fairness and incentives. CESifo working paper no. 1215.

Ferreira, D. (2002), Group loyalty and incentive pay. Working paper.

Hamilton, B.H.; Nickerson, J.A.; and H. Owan (2003): Team incentives and workerheterogeneity: an empirical analysis of the impact of teams on productivity and participation", *Journal of Political Economy*, vol. 111, issue 3, 465-498.

Hart, O. and Holmström, B. (1987), The theory of contracts. In T. F. Bewley (ed.): *Advances in economic theory.* Fifth World Congress. Cambridge: Cambridge University Press.

Holmström, B. (1982), Moral hazard in teams. *Bell Journal of Economics*, 13, 324-340.

Kandel, E. and Lazear, E. P. (1992), Peer pressure and partnerships. *Journal of Political Economy,* 100, 801-817.

Lazear, E. P. (1998), *Personnel economics for managers.* New York: Wiley.

Lazear, E. P. (2000), Performance pay and productivity. *American Economic Review*, 90, 1346-1361.

Marino, A. M. and Zabojnik, J. (2001), Profit centers and incentives in teams. Working paper.

Milgrom, P., and J. Roberts (1992): *Economics, Organization, and Management.* Englewood Cliffs, NJ: Prentice Hall.

Muthoo, A. (1999), *Bargaining theory with applications*. Cambridge: Cambridge University Press.

Paarsch, H. and Shearer, B. (2000), Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review*, 41, 59-92.

Rubinstein, A. (1982), Perfect equilibrium in a bargaining model. *Econometrica,* 50, 97-110.

Shearer, B. (2004), Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies*, 71, 513-534.

Strand, J. (2004), Equilibrium wage systems with moral hazard and monitoring costs. Working paper, Department of Economics, University of Oslo.

Towry, K. L. (2003), Control in a teamwork environment - the impact of social ties on the effectiveness of mutual monitoring contracts. *The Accounting Review*, 78, 1069-1095.