

## Social norms and economic incentives in firms

STEFFEN HUCK\*, DOROTHEA KÜBLER AND JÖRGEN WEIBULL

5 March 2003.

Revised version of IUI WP 565, 2001 ([www.iui.se](http://www.iui.se)).

**ABSTRACT.** This paper studies the interplay between economic incentives and social norms in firms. Our main focus is on “efficiency norms,” that is, norms that arise from workers’ desire for, or peer pressure towards, social efficiency for the workers as a group. In a simple model of team production and linear contracts we examine the possibility of multiple equilibria, and whether too strong economic incentives may knock out more efficient norm-based equilibria. We also consider partnerships, complementarity in production, stochastic production, heterogeneity among workers, and alternative incentive schemes including relative-performance pay.

*JEL code:* D23

*Keywords:* social norms, incentives, contracts.

### 1. INTRODUCTION

Empirical studies suggest that group norms in firms have important effects on behavior. Encinosa, Gaynor, and Rebitzer (1997) find that group norms matter in medical partnerships. Knez and Simester (1999) provide evidence for the airline industry and Ichino and Maggi (2000) for the banking industry.<sup>1</sup> For what follows, Ichino and Maggi’s finding is of particular interest: They report substantial shirking differentials between branches of a large Italian bank, despite identical monetary incentives

---

\*Huck is at University College London, Department of Economics & ELSE, Gower Street, London WC1E 6BT, United Kingdom; Kübler at Humboldt University Berlin, Department of Economics, Spandauer Str. 1, D-10178 Berlin, Germany; and Weibull at Department of Economics, Boston University, 270 Bay State Road, Boston MA 02115, USA. The authors are grateful for comments from Cedric Argenton, Tore Ellingsen, Armin Falk, Ernst Fehr, Jacob Glazer, Ed Lazear, Bentley MacLeod, Pedro Rey Biel, and from participants in presentations at Boston University, the Institute for Advanced Studies in Berlin, the Institute for Industrial Economics in Stockholm, the Stockholm Institute for Transition Economics, Wissenschaftszentrum Berlin (wzb), the CESifo Venice Summer Workshop 2002 and the ENTER Tilburg Jamboree 2003. This paper is an elaboration of the rough sketch in Weibull (1997).

<sup>1</sup>While Encinosa *et al.* focus on the interplay of group norms, multitasking and risk aversion, Knez and Simester show that firm-wide performance goals have an effect on employees who work in small groups, what allows them to monitor each other’s work effort closely.

governing the employees' efforts in these branches. They identify group-interaction effects as a key explanatory variable that allows for multiple equilibria.

This evidence is supplemented by experimental data. In a laboratory study, Falk, Fischbacher, and Gächter (2002) find that the same individual contributes more to a public good in a group with high average contributions than in a group with low contribution levels. Finally, Falk and Ichino (2003) report evidence on the effects of peer pressure in a recent non-laboratory experiment.

There are by now several papers in the economics literature where social norms have been included in microeconomic analyses,<sup>2</sup> but not many attempts have been made to study how social norms affect the incentive structure within firms.<sup>3</sup> Here we outline a new attempt in this direction and illustrate some of the consequences social norms can have for firm incentives and firm performance. We argue that social norms interact with economic incentives in a non-trivial but systematic way and may have a large effect upon firm performance. In this context, we take a social norm among the workers in a firm to be an informal ("unwritten") rule that is enforced either by the workers themselves or is internalized by these, but is not enforced by a third party, such as the employer. Once established in a group, a social norm is self-enforcing: expecting the others to adhere to the norm, each worker wants to adhere as well.

For a firm owner, social norms concerning work efforts are important because they affect profits. Suppose, for example, that a worker's compensation not only depends on his or her own effort but also on the effort of other workers. In the presence of such externalities, peer pressure, whether explicit or internalized, might penalize those who deviate from what is good for the group. And depending on the type of externality, which in its turn depends on the economic incentives in the firm, equilibrium output may be higher or lower than without the norm. On the one hand, if an increase in one worker's effort increases other workers' (expected) income, as in team production, a social norm may induce high efforts. On the other hand, if an increase in one worker's effort reduces others' (expected) income, such as under relative-performance pay schemes or under piece-rate schemes that are adjusted according to past performance, social norms may instead keep workers back from working hard.

---

<sup>2</sup>See e.g. Akerlof (1980), Moffitt (1983), Besley and Coate (1992), Bernheim (1994), Lindbeck, Nyberg and Weibull (1999, 2002), Hart (2001), and Kübler (2001), and the literature cited in these studies.

<sup>3</sup>For exceptions, see Kandel and Lazear (1992) and Barron and Gjerde (1997). Hart (2001) also focuses on norms and firms, but rather deals with the question whether the degree of trust between agents influences the optimal ownership structure. Also related is recent work by Rey Biel (2002) who studies how inequity aversion of agents affects optimal contracts.

We first analyze group norms in a simple model of team incentives. Each individual worker’s effort level is unobserved by the firm owner, but total output can be observed and verified without error.<sup>4</sup> We study the following basic scenario: First, the principal, who is the owner of the firm, chooses a base wage paid to all workers and a bonus proportional to the team’s output. Then, the agents, that is, the workers in the team, simultaneously choose their individual effort levels. We study the effect of a social norm concerning work effort among these team members. In particular, we show how the optimal incentive scheme, within the given class of incentive schemes, depends on the social norm.

A key observation is that social norms may cause multiplicity of equilibria under a given contract. Intuitively, multiplicity arises when a social norm introduces a coordination problem into the agents’ effort choices: others’ high or low efforts may serve as a “norm” for the individual worker. Kandel and Lazear (1992) develop a model of this phenomenon. However, they rule out multiplicity by assuming peer pressure to have certain convexity properties. By contrast, we here allow for multiplicity of equilibria since we believe that this explains some of the above-mentioned empirical evidence. Relaxing the Kandel and Lazear convexity condition, low-effort equilibria may coexist with high-effort equilibria — under the same incentive scheme. Such multiplicity is relevant for a principal who strives to maximize profits. For example, a firm in a high-effort equilibrium may suddenly “jump down” to a low-effort *and* low-profit equilibrium even when the incentive scheme is changed only slightly. Such an unwelcome surprise to a principal, who just increased workers’ economic incentives to induce them to work harder, can occur if the high-effort equilibrium, at which the firm was before the change, disappears in a fold of the equilibrium correspondence. In our view this sheds new light on so-called “crowding effects” of economic incentives, as discussed, for example, in Frey (1997) or Frey and Jegen (2003).<sup>5</sup>

From a dynamic perspective, it may thus be beneficial for a firm owner to temporarily change the economic incentive scheme, or in some other way inspire the workers to high efforts in order to lead the team of workers away from a low-effort equilibrium and help it coordinate on a “good” social work norm - with higher efforts, profit and worker utility. Once reached, such an equilibrium may remain intact even at less costly economic incentives, thanks to the newly created more demanding

---

<sup>4</sup>For the pioneering model of team production, see Holmstrom (1982).

<sup>5</sup>Drawing on the sociology and psychology literature Frey argues that economic incentives can crowd out *intrinsic motivation*. An early example for this effect goes back to Titmuss (1970) who argues that monetary incentives for blood donations undermine people’s intrinsic willingness to give blood. In contrast, the argument here is that economic incentives can weaken the effect of a *social norm*. Empirically, the two mechanisms might sometimes be hard to distinguish. However, our simple model offers diverse comparative static predictions that are, in principle, testable.

work norm.<sup>6</sup> Under the team incentive scheme in our model this is in everybody's interest. The equilibria associated with a given incentive contract are Pareto ranked according to effort: also the workers prefer high-effort to low-effort equilibria, given the economic incentives.

We also briefly analyze social norms under incentive schemes based upon relative performance. More specifically, we focus on the simplest possible case, namely when a worker's wage depends linearly on his relative effort - now assuming that individual workers' efforts - or some proxy for these - are observed by the principal. Like in the basic model, we assume that the principal first chooses a base wage paid to all workers and a bonus proportional to the relative performance of a worker, after which the workers simultaneously choose their individual effort levels. We again assume that the workers, besides their preference for income and leisure, have a preference to do what is good for the group of workers as a whole. In this environment, it is better for the group of workers if all reduce their efforts, while individual economic incentives push efforts up - in order to earn a higher wage. Again there may be multiple equilibria, associated with one and the same incentive scheme, but now these are not Pareto ranked - the owner and the collective of workers have diametrically opposed interests, given any bonus: while the owner prefers equilibria with high efforts, the workers prefer those with low efforts.

Our analysis of the team incentive case is related to recent independent work by Kandori (2002) on the dynamics of social norms in a public-goods provision environment. In contrast to our analysis, economic incentives are exogenous in his model, and the social norm in any time period depends on agents' behavior in the previous period. Applying the notion of stochastic stability (Kandori, Mailath, and Rob 1993, Young 1993) he finds that there is a range of stable effort levels, beginning at the (low) equilibrium effort without a norm and ending at a somewhat higher, and therefore socially more efficient, level. As a consequence of this multiplicity, situations with identical incentives may exhibit different norms and effort levels. Like in our model, and as in Kandori and Lazear (1992), the agents' utility in Kandori (2002) is additively separable into a material payoff, which depends on total effort and one's own effort, and a "psychological" payoff, which depends on one's deviation from others' behaviors. More precisely, the social norm in Kandori's model is given by the median effort in the group while in our model the norm is given by the average group effort (we also briefly study "conformity norms" which are closely related to the social norm analyzed in Kandori's model).

The remainder of the paper is organized as follows. The basic model of linear

---

<sup>6</sup>See Lindbeck, Nyberg, and Weibull (1999, 2002) for similar arguments in the context of transfers and taxes in a welfare state.

team incentives is introduced in Section 2, where we start by outlining the benchmark case of no social norm, and then point out some of the model's general qualitative properties. These general observations are illustrated in some detail in a numerical example given towards the end of the section. Section 3 briefly discusses some variations and generalizations of the basic model, while Section 4 discusses an alternative incentive scheme, a special case of relative performance pay. Section 5 concludes.

## 2. THE BASIC MODEL

Consider production in a firm with a profit-maximizing owner (the principal) as residual claimant. There are  $n > 1$  identical workers (agents). Each worker  $i$  exerts some effort  $x_i \geq 0$ . Let  $\bar{x}_{-i}$  denote the average effort exerted by all workers other than  $i$ , that is,  $\bar{x}_{-i} = \sum_{j \neq i} x_j / (n - 1)$ . The production technology is linear, and output  $y$  equals the sum of all workers' efforts,  $y = \sum_{i=1}^n x_i$ . The principal can only observe aggregate output  $y$ , not individual efforts. The workers can also observe aggregate output, and may or may not observe each others' individual efforts. From observing aggregate output  $y$ , each worker  $i$  can deduce the average effort  $\bar{x}_{-i}$  of the others (assuming knowledge of the production technology and his or her own effort).

In order to focus on the role of social norms in the simplest possible setting, we restrict the analysis in this section on linear contracts. They are common in practice and permit a transparent and straightforward analysis. More precisely, each worker receives the same wage  $w$ , and this *per capita* salary is an affine function of the firm's output  $y$ :

$$w = a + by/n, \quad (1)$$

where both the *base wage*  $a$  and the *bonus rate*  $b$  are assumed to be nonnegative.<sup>7</sup>

Workers' utility functions are additively separable, linear-quadratic in income and effort, and they may contain a term representing social preferences:

$$u_i = a + by/n - \frac{1}{2}x_i^2 - v(x_i, \bar{x}_{-i}, b), \quad (2)$$

where  $v : \mathbb{R}^3 \rightarrow \mathbb{R}$  is a continuous function to be specified below.

The firm is a price taker in its product market. We normalize the market price to unity, so the firm's profit — the residual left to the owner — is simply

$$\pi = y - nw = (1 - b)y - na. \quad (3)$$

---

<sup>7</sup>The non-negativity of the base salary means, for example, that it is not possible for the owner effectively to sell the firm to the workers. This nonnegativity assumption can be justified by wealth constraints or limited liability, see Section 3.7 for a brief discussion of contract forms, wealth constraints, limited and unlimited liability. The nonnegativity of the bonus rate will not be binding in equilibrium, see below.

The owner is assumed to be a risk neutral profit-maximizer.

The interaction takes the form of a two-stage game, where the owner first chooses a contract  $(a, b)$ , and then all workers observe the contract and simultaneously choose their individual efforts  $x_i$ . Hence, a strategy for the owner is a vector  $(a, b) \in \mathbb{R}_+^2$ , and a strategy for a worker  $i$  is a function, or “rule,”  $s_i$  that assigns an effort level  $x_i = s_i(a, b) \in \mathbb{R}_+$  to every contract  $(a, b)$ . We solve this game for subgame-perfect equilibrium.

More precisely, we will focus on *symmetric* equilibria, that is, equilibria in which all workers use the same strategy and, thus, exert the same effort under any given contract  $(a, b)$ . A uniform effort profile  $(x, x, \dots, x)$  constitutes a symmetric Nash equilibrium in the corresponding subgame if and only if the common effort  $x$  is a worker’s best reply when the others exert effort  $x$ :

$$x \in \arg \max_{t \geq 0} a + \frac{b}{n}t - \frac{1}{2}t^2 - v(t, x, b). \quad (4)$$

It is obvious that workers’ efforts do not depend on the base wage  $a$ . Hence, since such a salary is costly to the firm owner, we have  $a = 0$  in every subgame perfect equilibrium.<sup>8</sup>

If  $v$  is continuously differentiable in its first argument, as in some cases to be studied below, then a necessary condition for an interior symmetric subgame Nash equilibrium (one in which all workers exert the same *positive* effort) is that the above maximand’s derivative with respect to  $t$ , the worker’s own effort, vanish at  $t = x$ . This is equivalent to

$$x = \frac{b}{n} - v'_1(x, x, b). \quad (5)$$

From this we see that social preferences may play a role in determining the set of equilibria. Below we will discuss norms that are driven by social conformity and social efficiency, respectively. In some of these cases, there can even exist multiple (symmetric) equilibrium efforts for one and the same bonus rate.

What then is a social norm in this context? We shall say that a *social norm* is a rule prescribing a certain effort level, depending on what effort levels others (are expected to) choose. The social norm is thus endogenous. An equilibrium effort  $x$  is the effort “dictated” by the social norm when the others exert the same effort  $x$ , while the *social preferences*, here represented by the function  $v$ , in their interplay with the economic incentives of the workers, determine which effort levels  $x$  are equilibria. This is the terminology that we will use most of the time. However, we shall sometimes be slightly more lax and refer to the absence of social preferences as the model “without a social norm.”

---

<sup>8</sup>There are Nash equilibria with strictly positive base wages but these are not subgame perfect.

**2.1. No social preferences.** In order to obtain a benchmark for the subsequent analysis, we first study the case when workers' preferences are not social, that is, when  $v \equiv 0$ . From (2) it is immediate that workers' decisions concerning effort are strategically independent in this special case. Regardless of whether workers decide simultaneously or sequentially, each worker just solves

$$\max_{x_i \geq 0} \frac{b}{n} x_i - \frac{1}{2} x_i^2. \quad (6)$$

Consequently, the unique Nash equilibrium effort level, given any contract  $(a, b)$ , is  $x_i = b/n$  for all workers  $i$ .

We solve for subgame-perfect equilibrium by inserting the equilibrium effort into the expression for the firm's profit, and obtain

$$\pi = (1 - b)b - na. \quad (7)$$

Since this is a decreasing function of the base wage, the owner will, as noted in general above, set  $a = 0$ , and the owner's choice of bonus rate is simple:  $b = 1/2$ .

In sum: under linear technology and in the absence of a social norm, there exists a unique subgame-perfect equilibrium. In this equilibrium, the owner offers zero base salary, and offers a 50/50 split of the firms' revenue with the team of workers. Workers' effort level on the unique subgame-perfect equilibrium path is  $x = \frac{1}{2n}$ .

Note also the free-riding among workers. Given any contract, their common equilibrium effort is  $x = b/n$ , while the sum of their utilities achieves its unique maximum at  $x = b$ :

$$\sum_i u_i = n \left( a + bx - \frac{1}{2} x^2 \right). \quad (8)$$

Hence, if they could, the workers as a collective would like to commit to the much higher effort level  $x = b$ , given any bonus rate  $b$ . We will call this effort level the *team optimum* effort for the workers as a collective under contract  $(a, b)$ , and we will sometimes denote this  $\hat{x}(a, b) = b$ .

Interestingly enough, the owner would offer the same contract also if workers could commit to a common effort level in this way (granted the owner knew they will commit to a common effort level). For the workers would then choose  $x = b$  for all  $b$ , resulting in profits  $\pi = n(1 - b)b - na$ , so the owner would choose  $a = 0$  and  $b = 1/2$ . In other words, then the owner would still offer the same contract, and everyone, including the owner, would be better off than without commitment. However, such commitments are rarely available in practice, so we address the question if instead a social norm can induce workers without commitment power to exert equilibrium efforts closer to the team optimum.

**2.2. Conformity preferences.** The simplest kind of social norm is when each worker, *ceteris paribus*, prefers to exert the same effort as the others. In other words, each worker experiences disutility from “sticking out” in terms of work effort. This amounts to a social norm to “do like others.” Such a *preference for conformity* can be represented in the present model as the special case when  $v(x_i, \bar{x}_{-i}, b)$ , for any given values of  $\bar{x}_{-i}$  and  $b$ , achieves its minimum value at  $x_i = \bar{x}_{-i}$ . Consider, for instance,  $v(x_i, \bar{x}_{-i}, b) = c(x_i - \bar{x}_{-i})$  where  $c$  is some convex function with minimum at zero.<sup>9</sup>

A subutility function  $v$  with this general qualitative property is consistent with an internalized social norm under which the individual worker has a preference for doing what others do, irrespective of whether other workers can or cannot observe his or her efforts. By contrast, under an externally enforced norm, that is, if social utility arises only from avoiding others’ “punishment” of defectors — by physical, economic or social means, or maybe only by a frown — then the present utility representation requires that workers can observe each others’ efforts, or at least some signal correlated with these.<sup>10</sup>

However, if the function  $v$  is continuously differentiable, then the necessary first-order condition (5) for symmetric equilibrium becomes identical with the one that is obtained in the absence of a social norm. For we then necessarily have  $v'_1(x, x, b) = 0$ , since  $v(t, x, b)$ , viewed as a function of the worker’s own effort  $t$ , is minimized at  $t = x$ . Hence, a preference for conformity, with continuously differentiable utility, does not help to explain other (symmetric) equilibrium behaviors than those generated by the benchmark model without social preferences.

However, a conformity norm may have an effect even on symmetric equilibrium with identical workers if the disutility function is non-differentiable “on the diagonal,” that is, where all workers exert equal effort.<sup>11</sup> For example, suppose that

$$v(x_i, \bar{x}_{-i}, b) = \gamma |x_i - \bar{x}_{-i}| \tag{9}$$

for some  $\gamma \in (0, b/n)$ . Then every effort level in the interval  $[b/n - \gamma, b/n + \gamma]$  is the common equilibrium effort level in some symmetric Nash equilibrium.<sup>12</sup> Every such equilibrium effort level is a potential social norm—a behavior that is self-enforcing once others (are expected to) do likewise. Unlike in the differentiable case, a worker’s

---

<sup>9</sup>Kandel and Lazear (1992) develop a model of this type, op. cit. p.811.

<sup>10</sup>As mentioned before, we do not model the incentives to execute such punishments.

<sup>11</sup>A rationale for even a discontinuity when all agents take the same action is given in Bernheim (1994), where even the slightest deviation from others’ behavior in equilibrium is interpreted as the act of an “extremist,” causing a finite fall in the deviator’s social esteem, which enters his utility function in Bernheim’s model.

<sup>12</sup>To see this, note that  $i$ ’s utility is maximized at  $x_i = \bar{x}_{-i}$  iff  $b/n - \bar{x}_{-i} + \gamma \geq 0$  and  $b/n - \bar{x}_{-i} - \gamma \leq 0$ . Hence,  $b/n - \gamma \leq x \leq b/n + \gamma$  in symmetric equilibrium.

marginal disutility from norm defection, at the norm effort level, is no longer zero. This difference from the differentiable case is enough to prevent small deviations from effort norms in the above range of common effort levels, containing the benchmark level  $x = b/n$ .<sup>13</sup> Any symmetric equilibrium effort  $x$  can be viewed as a social norm in the above-mentioned sense: all individuals are expected to exert this effort, and, given such expectations, this effort indeed is the unique optimal effort for every worker.

We believe, however, that social preferences are not only concerned with "doing like others," but also with "doing what is good for one's peer group," and that the latter motive is stronger the more others in the peer group do what is good for the peer group. We elaborate on such social preferences in the next subsection, assuming that the relevant peer group of each worker is the team of workers in the firm.<sup>14</sup>

**2.3. Efficiency preferences.** In the benchmark model, each agent's efforts have positive externalities for the other agents: If  $i$  works harder, all other workers benefit by way of earning a higher income. Clearly, whenever actions cause externalities on others, self interest and group interest may conflict. We assume here that each worker  $i$  has social preferences with the following qualitative feature: if others cause higher positive externalities, or worker  $i$  expects them to do so, then also  $i$  wants to work more, *ceteris paribus*. This is the idea of *peer pressure*. Such peer pressure can be internalized or external: arising in the mind of the defector (say, by way of reduced self-esteem), and/or arising from other team members' irritation or anger from losing income due to  $i$ 's shirking. Here, we effectively study an internalized social norm.<sup>15</sup>

Assume, thus, that workers may derive disutility when deviating from the team optimum effort level, and more so, the harder others in the team work. We formalize such preferences by assuming that the social disutility term  $v(x_i, \bar{x}_{-i}, b)$  is convex in the worker's own effort, with minimum at  $x_i = \hat{x}(a, b) = b$ , the team optimum effort level, and non-decreasing in  $\bar{x}_{-i}$ , other workers' average effort.<sup>16</sup>

---

<sup>13</sup>In a similar model, Kandori (2002) also obtains an interval of effort levels. We note two differences. First, in his model, the norm is expressed by the median effort, and not as here by the average effort of others. (Thus, he has to assume that workers can observe each others' individual effort levels.) Second, Kandori assumes that individuals suffer no disutility when exceeding the norm, and consequently his equilibrium interval has the benchmark effort as its lower bound, and not, as here, as its midpoint.

<sup>14</sup>An interesting avenue for research is to study the formation of peer groups.

<sup>15</sup>For simplicity, we do not model the incentive problem arising when defectors have to be punished. Interestingly, Fehr, Sigmund and Nowak (2002) find that many individuals have a preference for punishing deviators, even at a cost to themselves. This would reduce or even solve the incentive problem of norm enforcement.

<sup>16</sup>Indeed, the qualitative properties to be demonstrated are unrelated to the team optimum -

Under these assumptions, it follows from standard arguments that there exists at least one symmetric Nash equilibrium under every contract  $(a, b)$ , and the common equilibrium effort level  $x$  in any such equilibrium satisfies  $b/n \leq x \leq b$ . Hence, workers exert more effort than in the benchmark model without a social norm, but they do exert less than the worker optimum effort. Formally, and somewhat more generally:

**Proposition 1.** *Suppose that  $v : \mathbb{R}_+^3 \rightarrow \mathbb{R}$  is continuous, and that  $v(t, x, b)$ , for any  $b, x > 0$ , is convex in  $t$  with minimum at  $t = b$ . Then there exists at least one symmetric Nash equilibrium effort  $x$  for each  $b > 0$ , and all such  $x$  lie in the interval  $[b/n, b]$ .*

**Proof** Suppose  $b > 0$ . A common effort level  $x$  is a Nash equilibrium associated with contract  $(a, b)$  if and only if it satisfies (4). First note that, for each  $x \geq 0$ , the set  $\xi(x)$  on the right-hand side in (4) is a nonempty, convex and compact subset of the interval  $[b/n, b]$ . No  $t$  outside this interval belongs to the set  $\xi(x)$ , since the maximand is strictly increasing in  $t$  to the left of this interval, and it is strictly decreasing to the right of the interval. Hence,  $\xi(x)$  is a nonempty and compact subset of  $[b/n, b]$ , by Weierstrass' Maximum Theorem applied to the continuous maximand. Moreover, if  $v$  is convex, then the maximand is concave, and thus  $\xi(x)$  is convex. Second, the so-defined solution correspondence  $\xi$ , which maps others' mean effort  $x$  to own optimal efforts  $t \in \xi(x)$ , is upper hemi-continuous by Berge's Maximum Theorem. Being a compact- and convex-valued u.h.c. correspondence from  $\mathbb{R}_+$  to  $[b/n, b] \subset \mathbb{R}_+$ ,  $\xi$  has at least one fixed point  $x$  in  $[b/n, b]$ , by Kakutani's Fixed-Point Theorem.  $\square$

For any bonus rate  $b > 0$ , let  $X^{NE}(b) \subset [b/n, b]$  be the associated non-empty set of (uniform) Nash equilibrium effort levels, the solutions to (4). If  $v$  is continuously differentiable, then the first-order condition (5) is both necessary and sufficient for symmetric and interior subgame Nash equilibrium, and we necessarily have  $x < b$ , since  $v'_1 = 0$  when its first argument, the worker's own effort, takes the value  $b$ . This is the case, for example, if  $v$  is proportional to the square of the worker's deviation from the socially optimal effort:

$$v(x_i, \bar{x}_{-i}, b) = \frac{1}{2} (\hat{x}(a, b) - x_i)^2 p(\hat{x}(a, b) - \bar{x}_{-i}) \quad (10)$$

where, in the simple case considered here,  $\hat{x}(a, b) = b$ , and where  $p : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is continuously differentiable with  $p' \leq 0$ . Hence,

---

what matters is that the social disutility term is non-increasing in own effort in some range above the benchmark equilibrium effort  $b/n$ .

$$v(x_i, \bar{x}_{-i}, b) = \frac{1}{2} (b - x_i)^2 p (b - \bar{x}_{-i}). \quad (11)$$

We will refer to the proportionality factor  $p(b - \bar{x}_{-i})$  as *peer pressure*. The more others deviate from the social optimum, the less peer pressure does a deviating worker feel, *ceteris paribus*.

In this special case, the first-order condition (5) becomes  $x = F(x)$ , where

$$F(x) = \frac{1 + np(b - x)}{1 + p(b - x)} \cdot \frac{b}{n}. \quad (12)$$

The function  $F : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is clearly continuous and non-decreasing, with  $F(0) \geq b/n$  and  $F(b) < b$ . Hence, we can immediately verify that there indeed exists at least one fixed-point under  $F$ , and that all fixed points belong to the interval  $[b/n, b)$ , where every fixed point is the effort level in a symmetric Nash equilibrium and *vice versa*.

The above machinery allows us to characterize subgame perfect equilibrium in a fairly concise manner. First, since workers' efforts are independent of the base salary  $a$ , and this is costly to the firm owner,  $a = 0$  in any subgame perfect equilibrium. Second, we have  $\pi = (1 - b)nx$  in all symmetric subgame equilibria. Hence, iso-profit curves in the  $(b, x)$  plane are hyperbolas.<sup>17</sup> Moreover, a bonus-effort pair  $(b, x)$  is the *outcome of a symmetric subgame perfect equilibrium* if and only if (i)  $x \in X^{NE}(b)$ , and (ii) for every bonus rate  $b' \neq b$  there exists a  $x' \in X^{NE}(b')$  such that the associated profit  $\pi' = (1 - b')nx$  does not exceed the profit  $\pi = (1 - b)nx$  under  $(b, x)$ .

We finally note that if the peer-pressure function  $p$  is twice differentiable and concave, then so is  $F$ , and hence the equilibrium effort is unique:

**Proposition 2.** *If  $p$  is twice differentiable, and  $2[p']^2 \geq [1 + p]p''$  for all  $x \in (0, b)$ , then the fixed point  $x$  under  $F$  is unique.*

**Proof** The uniqueness claim follows immediately from straight-forward differentiation:

$$F''(x) = b(1/n - 1) \left( 2[p'_1]^2 - [1 + p]p''_{11} \right) / [1 + p]^3,$$

and hence  $F$  is concave if  $2[p'_1]^2 \geq [1 + p]p''_{11}$ .  $\square$

Before proceeding to extensions and variations of this model, we illustrate it by way of a numerical example.

---

<sup>17</sup>For any profit level  $\pi$ ,  $x = \pi [n(1 - b)]^{-1}$ .

**2.4. Example.** Suppose that the peer-pressure function  $p$  is given by

$$p(b-x) = \begin{cases} \alpha \exp(-\beta(b-x)^2) & \text{if } x \leq b \\ \alpha & \text{if } x > b \end{cases} . \quad (13)$$

for  $\alpha, \beta > 0$ . Here,  $\alpha$  represents the utility weight placed on norm adherence and  $\beta$  the sensitivity of this weight to others' norm adherence. Figure 1 shows the graph of the associated function  $F$  for  $n = 5$ ,  $b = 0.45$ ,  $\alpha = 4$  and  $\beta = 40$ . The diagram shows three equilibria - these are the intersections of the graph of  $F$  with the diagonal. Only the high- and low-effort equilibria are stable under adaptive dynamics; a small deviation (up, down) from the medium-effort equilibrium induces a movement (up, down) towards the high- or low-effort equilibrium level. We will therefore largely neglect the medium-effort equilibrium in much of the subsequent discussion. We also note that all three equilibrium effort levels lie - as they should - between the social-preference free Nash equilibrium level  $b/n = 0.09$  and the team optimum level  $b = 0.45$ .

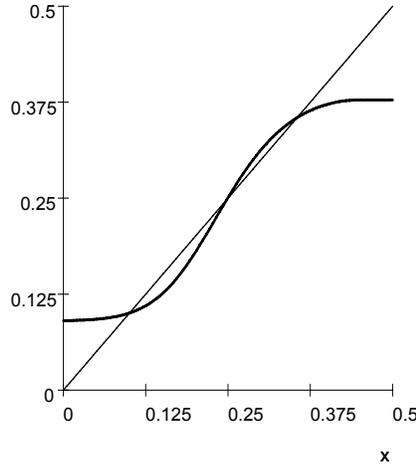


Figure 1: The Nash equilibrium fixed-point equation  $x = F(x)$ , when the bonus rate  $b$  is 0.7.

Figure 2 shows the Nash equilibrium manifold — the set of Nash equilibrium effort levels  $x$ , on the vertical axis, associated with each bonus rate  $b$ , on the horizontal axis (generated for the same parameter values as used in Figure 1). The "inverted S" curve is the graph of the Nash equilibrium correspondence: for each bonus rate  $b$

it provides the set of effort levels  $x$  such that the effort profile  $(x, x, \dots, x)$  is a Nash equilibrium in the associated subgame. The thin steep straight line represents the team optimum effort level,  $x = b$ , and the less steep dotted straight line the Nash equilibrium effort level in the absence of social preferences,  $x = b/n$ . The hyperbolas are iso-profit curves.<sup>18</sup> The lowest iso-profit curve corresponds to the profit level in the benchmark Nash equilibrium, obtained at  $b = 1/2$ , and the highest iso-profit curve corresponds to the profit level in case all workers could collectively commit to a common effort level. The resulting profit is the highest possible, and, as noted in general, also this profit maximum is obtained at  $b = 1/2$ . The intermediate iso-profit curve corresponds to the highest equilibrium profit achievable when workers have social preferences of the form assumed here. This profit maximum is achieved at  $b^* \approx 0.44$ .

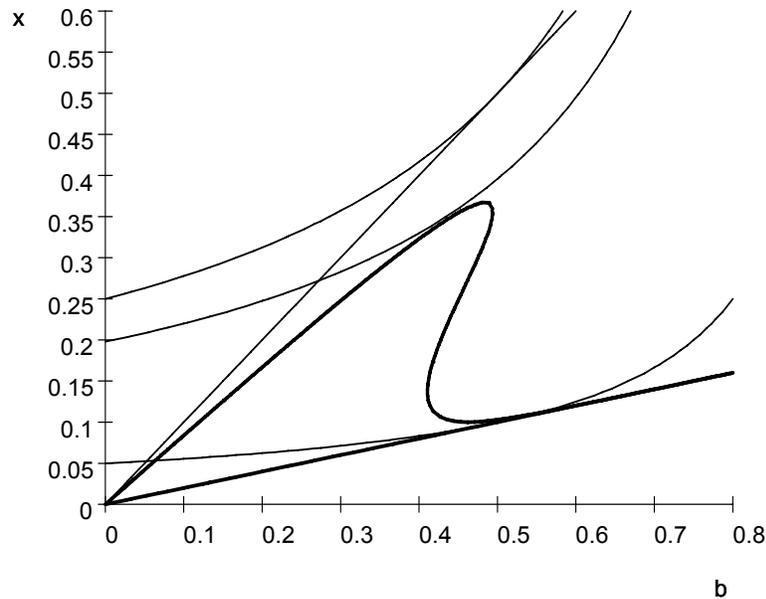


Figure 2: Nash equilibrium efforts  $x$  (vertical axis) for different bonus rates  $b$  (horizontal axis).

The diagram also shows that, for all bonus rates  $b$  below a certain critical value,  $b_1 \approx 0.41$ , the associated subgame Nash equilibrium effort is unique. Moreover, in this interval,  $(0, b_1)$ , the equilibrium effort is increasing in the bonus rate  $b$ . At the

<sup>18</sup>Each iso-profit curve is of the form  $b = 1 - c/x$ , for  $c = \pi/n$ .

critical value  $b_1$  there are two equilibria. Then follows an interval,  $(b_1, b_2)$  with three equilibria; with high, medium and low effort, respectively. At the second critical bonus rate,  $b_2 \approx 0.48$ , there are again only two equilibria, and for all higher bonus rates the equilibrium effort is again unique. In particular, the *first-best* bonus rate, that is the rate  $b^* \approx 0.44$  that maximizes profits across the whole subgame Nash equilibrium manifold, though being subgame perfect (see below), entails multiple equilibria. Thus, a principal who chooses the bonus rate in this way, in expectation of the corresponding high-effort equilibrium, might face a disappointing surprise. If workers instead play the low-effort (or medium-effort) equilibrium at that bonus rate, his profits will be considerably lower than anticipated. By contrast, the owner may guarantee a higher equilibrium profit than in this worst-case equilibrium under the first-best bonus rate by instead choosing a slightly lower bonus rate - below the first critical bonus rate  $b_1$  - where the associated subgame equilibrium is unique, say at  $b = 0.4$ . However, no such bonus rate is consistent with subgame perfection.

Indeed, what bonus-effort pairs  $(b, x)$  are consistent with subgame perfection in this example? Clearly  $(b^*, x^*)$ , where  $x^* \in X^{NE}(b^*)$  is the "high effort" Nash equilibrium, is such. No other bonus rate has an accompanying subgame equilibrium effort level resulting in a higher profit. However, there are infinitely many symmetric subgame perfect equilibrium outcomes  $(b, x)$  in this example. These are precisely those pairs  $(b, x)$  such that  $b_1 \leq b \leq b_2$  and  $x \in X^{NE}(b)$  is either the high or the medium effort.<sup>19</sup> In other words, in this example, multiplicity of subgame Nash equilibria is unavoidable in symmetric subgame-perfect equilibrium.

Given a bonus rate for which there exist multiple symmetric subgame Nash equilibria, which of these is more likely to prevail? One potential avenue for such an investigation is offered by evolutionary game theory. The interaction between the workers constitutes a coordination game, and the tools developed by Kandori, Mailath, and Rob (1993), Young (1993), and Benaim and Weibull (2002), among others, could be applied to a discrete-choice representation of the subgame. However, such an analysis falls outside the scope of the present study. We note in passing, though, that the high-effort equilibrium, as noted above, Pareto dominates the low-effort equilibrium.<sup>20</sup> Hence, both (product) market selection and worker mobility speak against the low-effort equilibrium: a lower profit makes the firm's survival less likely, and a lower utility to the workers makes them less willing to keep their jobs. Indeed, if workers in the relevant labor market have a reservation utility above that in a low-

---

<sup>19</sup>However, recall that we would discard all "medium effort" equilibria since they are unstable in adaptive dynamics.

<sup>20</sup>It would be an interesting game-theoretic exercise to study whether the high effort equilibrium risk-dominates the low effort equilibrium, in which case the high-effort equilibrium would be a strong candidate for evolutionary selection.

effort equilibrium but below that in a high-effort equilibrium, and if transaction costs are low, then workers will leave a firm which is trapped in the low-effort equilibrium. Actually, if workers have reservation utilities between the low-effort and high-effort equilibrium utilities, then also *forward induction* arguments speak for the high-effort equilibrium.<sup>21</sup>

We also note that if there is some inertia in workers' effort adjustment to changes in the bonus rate, and if we disregard the (potentially unstable) medium-effort equilibrium, then the economist's standard intuition holds for most bonus rates: a small increase in the bonus rate unambiguously leads to an *increase* in effort. If the bonus is close to  $b_2$ , however, a small increase in the bonus rate may lead to a sudden large fall in effort, as the team of workers "falls down" to the new and unique low-effort equilibrium.

A marginally increased bonus may thus have a drastic adverse effect on the equilibrium effort level. The effectiveness of the social norm is destroyed when incentives become too strong.<sup>22</sup> This potentially negative effect of increased economic incentives can also be explained by studying how a worker's best reply to the others' average effort changes as the bonus rate changes for social disutility functions of the form (11). Keeping other workers' average effort  $\bar{x}_{-i}$  fixed, and assuming that worker  $i$ 's optimal effort  $x_i^*$  then is a differentiable function of the bonus rate, we obtain from (4) that

$$\frac{\partial x_i^*}{\partial b} = \frac{1/n}{p+1} + \frac{p}{p+1} + \left(1 - \frac{1}{n}\right) \frac{bp'}{(p+1)^2} \quad (14)$$

The first two terms are positive, and the third is either zero or negative, since  $p' \leq 0$ .<sup>23</sup> That is, if the slope  $|p'|$  of the peer pressure function is sufficiently large, then the direct effect of increasing the bonus is a *reduction* of agent  $i$ 's effort. In adjustments towards equilibrium, this will cause reductions in the other agents' efforts as well, which will cause further reductions' in everybody's efforts and so on. In other words, when  $\partial x_i^*/\partial b$  turns negative, the subgame equilibrium effort level decreases as the bonus rate increases. This corresponds to a downward sloping segment of the Nash equilibrium manifold in Figure 2.

The intuition for this phenomenon is as follows. An increase in  $b$  has *three* effects on agent  $i$ 's effort (holding others' efforts constant). It increases  $i$ 's incentive to work

<sup>21</sup>See Asheim and Dufwenberg (2002) for a general game theoretic argument along these lines.

<sup>22</sup>Similarly, it may happen that a reduction in economic incentives induces a sudden increase of efforts. From the principal's point of view, this might be an unexpected but welcome surprise.

<sup>23</sup>Note also that the third term equals  $-F'(x)$ . The absolute value of the third term is thus less than 1 in case of a unique equilibrium, as well as in the low-effort and high-effort equilibria in both examples (but not in the middle-effort equilibrium).

hard via the monetary reward for higher output.<sup>24</sup> This is reflected by the first term in (14). It also increases  $i$ 's effort by increasing  $i$ 's deviation from what is optimal for the group. This is reflected by the second term in (14). However, at the same time an increase in  $b$  *reduces* agent  $i$ 's effort as the others' average deviation from the norm also increases (their efforts are fixed, by assumption), which reduces the peer pressure felt by worker  $i$ . And it is exactly this effect which is captured by the negative third term in (14). If the third effect outweighs the first two, agent  $i$  wants to reduce his efforts which, in the aggregate, causes a cascade of downward adjustments, and results in a new equilibrium with a lower common effort, and this new effort level may lie far below the effort level in the original equilibrium (such as at a fold).

In our view this analysis sheds new light on so-called crowding effects of economic incentives as discussed, for example, in Frey (1997). We offer an alternative explanation for this phenomenon based on social norms. Due to the fact that social pressure decreases when others fall short of the team optimum, the norm may lose power when the bonus is increased. Thus, monetary incentives may reduce social incentives through this social interaction effect whereas the typical crowding argument focuses on psychological mechanisms at the individual level.

### 3. VARIATIONS

In this section we briefly study various alternatives to and extensions of our model, in order to illuminate the robustness of the above qualitative results, and to see how they can be generalized.

**3.1. Partnerships.** A partnership, that is, a firm owned by workers who split the profit, which we here presume to be in equal shares, is equivalent to setting  $b = 1$  in our model. The worker optimum effort level thus becomes  $x = 1$ , while the Nash equilibrium effort in the case without social preferences is  $x = 1/n$ . The fixed-point equation that determines (symmetric) equilibrium effort levels in cases with social preferences has the same qualitative properties as in the basic model. In particular, for certain peer-pressure functions there exist multiple equilibria, and these are Pareto ranked according to their effort levels. As in Kandel and Lazear's (1992) analysis of peer pressure in partnerships, output is typically closer to the efficient level, due to such a work norm. Our model adds to their analysis by way of allowing for multiple equilibria.

---

<sup>24</sup>Under more general preferences, this need not be the case, since there could be a strong and negative income effect.

**3.2. Decreasing returns to scale.** How do the results depend on the linearity of the production function? We briefly consider the case of decreasing returns. More precisely, let  $y = f(L)$ , where  $L = \sum_j x_j$  and  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is twice differentiable, with  $f' > 0$ ,  $f'' < 0$ , and assume the disutility function  $v$  is of the form (11). The socially optimal effort level  $\hat{x}(a, b)$ , given any contract  $(a, b)$ , is uniquely determined by the first-order condition for social (team) efficiency:  $x = bf'(nx)$ . The right-hand side of this equation is continuous and decreasing in  $x$ , and hence it has a unique solution for every  $b$  and  $n$ . Going back to (10) the fixed-point equation (12) for symmetric (subgame) Nash equilibrium effort,  $x = F(x)$ , now has

$$F(x) = \frac{f'(nx) + np(\hat{x}(a, b) - x)}{1 + p(\hat{x}(a, b) - x)} \times \frac{b}{n}. \quad (15)$$

We note that the function  $F$  need no longer be increasing. In particular, if the production function  $f$  meets the usual Inada condition that  $f'(x) \rightarrow +\infty$  as  $x$  goes to zero, then also  $F(x) \rightarrow +\infty$  as  $x$  goes to zero.

**3.3. Complementarity in effort.** So far, all workers' efforts have been perfect substitutes. Consider now the extreme opposite case when individual efforts are perfect complements. Hence, suppose  $y = \min_j x_j$ . The team optimum level is now the common individual effort  $x$  which maximizes  $bx - nx^2/2$ . Thus,  $\hat{x}(a, b) = b/n$ .

As will be seen below, in the absence of the group norm, any effort level in the interval  $[0, b/n]$  is consistent with Nash equilibrium, and in the presence of the group norm, the set of equilibrium effort levels is a subinterval of the form  $[\lambda, b/n]$ , where  $\lambda > 0$  is larger the more important the social norm is in comparison with economic incentives (as measured by the peer pressure function  $p$ ). To see this, assume that all workers  $j \neq i$  choose the same effort level  $x$ , and let  $v$  be as in equation (11). Then the utility to worker  $i$  depends on  $i$ 's own effort in the following way:

$$u_i = \begin{cases} bx_i/n - \frac{1}{2}x_i^2 - \frac{1}{2}(b/n - x_i)^2 p(b/n - x) & \text{if } x_i \leq x \\ bx/n - \frac{1}{2}x_i^2 - \frac{1}{2}(b/n - x_i)^2 p(b/n - x) & \text{if } x_i > x \end{cases} \quad (16)$$

Hence, as a function of his own effort  $x_i$ , the utility to worker  $i$  is continuous, but has a kink precisely at the others' effort level  $x$ . It is easily verified that a necessary and sufficient condition for symmetric Nash equilibrium, where all workers choose the same effort  $x$ , is given by the following double inequality

$$\frac{p(b/n - x)}{1 + p(b/n - x)} \frac{b}{n} \leq x \leq \frac{b}{n}. \quad (17)$$

Let us now consider various intensities of the social norm by replacing the peer pressure function  $p$  with some  $\tilde{p} = \lambda p$ . We note, in particular, that in the absence of

the group norm, that is, when  $\lambda = 0$ , inequality (17) is met by every effort level  $x$  in the whole interval  $[0, b/n]$ , as claimed above. We also note that, as the utility weight  $\lambda$  placed on the peer pressure term increases, the interval of Nash equilibrium effort levels shrinks, and in the limit as  $\lambda \rightarrow +\infty$  becomes the singleton set  $\{b/n\}$ .

In sum, the effect of a preference for social efficiency, when workers' efforts are perfect complements, is to reduce the continuum set of equilibrium effort levels towards the team optimum effort.<sup>25</sup>

**3.4. Sequential effort choice.** While in our model workers choose effort levels simultaneously, in some real-life situation these choices are made sequentially, or at least in part sequentially (e.g. one worker takes the lead and the others follow). How does this affect the analysis?

For the sake of clarity, we consider the special case of only two workers. Both workers have utility functions as given by (2) and (11). Having observed the first worker's effort choice,  $x_1$ , the second worker chooses his or her own effort,  $x_2$ , in order to maximize

$$\frac{b}{2}x_2 - \frac{1}{2}x_2^2 - \frac{1}{2}(b - x_2)^2p(b - x_1). \quad (18)$$

Consequently, worker 2 will choose effort according to the reaction function (c.f. equation (12))

$$x_2 = F(x_1) = \frac{1 + 2p(b - x_1)}{1 + p(b - x_1)} \times \frac{b}{2}. \quad (19)$$

Differentiation w.r.t.  $x_1$  verifies our earlier claim that efforts are strategic complements:  $F' \geq 0$ , with  $F' > 0$  if  $p' < 0$ . Anticipating the second worker's reaction, the first worker chooses his or her effort,  $x_1$ , in order to maximize

$$\frac{b}{2}(x_1 + F(x_1)) - \frac{1}{2}x_1^2 - \frac{1}{2}(b - x_1)^2p(b - F(x_1)). \quad (20)$$

There are two countervailing effects determining the optimal choice. On the one hand, the first worker has an additional incentive to work hard, beyond his incentive in a simultaneous-choice situation. The reason is that working more increases the second worker's effort, via the group norm. The first worker profits from this higher effort by the marginal utility gain  $bF(x_1)/2$ . On the other hand, the first worker can also reduce the peer pressure on himself. If he works less, the second worker will also

---

<sup>25</sup>Note that also in a more general model with heterogeneous agents, first best effort of team members can be implemented as a Nash equilibrium when production is characterized by perfect complementarity of inputs. See Legros and Matthews (1993) and Vislie (1994). However, the efficient effort levels do not constitute a unique equilibrium under these schemes, again leaving room for social norms to limit the set of possible equilibria.

work less, and the value of the peer-pressure function  $p$  will fall. Which of these two effects dominates will, in general, depend on the exact shape of  $p$ . If  $p$  is not too steep, the first effect will dominate. Then both workers' efforts will be greater with sequential moves than with simultaneous moves. Hence, in such situations it is in the firm owner's interest to organize work accordingly.<sup>26</sup>

**3.5. Stochastic production.** Until now we have assumed that production is deterministic. But the model can readily be generalized by adding a stochastic element to the production technology. In the absence of a group norm this is entirely straightforward. In the presence of a group norm, a stochastic relation between input and output raises a question about the underlying information structure—of what agents can observe.

To be more specific, let individual output be  $y_i = x_i + \varepsilon_i$ , where the  $\varepsilon_i$ 's are identically and statistically independently distributed random variables, and let the model be otherwise as in section 2. In particular, team output is  $y = \sum_i y_i$ , and all workers and the owner are risk neutral. Introducing risk aversion would not change the qualitative effects of a social norm as long as we keep the assumption of additive separability in wage, effort cost and the effect of the social norm. In the absence of a norm (see Section 2.1), the analysis remains unchanged.<sup>27</sup> Suppose each worker only knows his or her own effort ( $x_i$ ) and output ( $y_i$ ) as well as total output ( $y$ ) but not the individual efforts of the other workers (beyond inferring  $y_{-i}$  from observation of  $y$  and  $y_i$ ). In this setting we need to modify the definitions (2) and (10) of the utility to a worker.

First consider the case in which the norm is based on effort, not output. In this case, judgement is not based on consequences but on the action taken. A worker does not feel bad if she knows that she exerted the right effort but was unfortunate and had a negative shock to her output  $y_i$ . More specifically, the strength of the social norm now depends on one's own deviation from the ideal effort but only indirectly on the others' (unobserved) actual deviations, through one's observation of their joint output,  $y_{-i} = y - y_i$ . In this case, an adaptation of (10) suggests

$$u_i = w(y) - \frac{1}{2}x_i^2 - \frac{1}{2}(\hat{x} - x_i)^2 p(\hat{y}_{-i} - \bar{y}_{-i}) , \quad (21)$$

where  $\hat{x}$  is the individual effort which maximizes the sum of all workers' expected utility from consumption and effort, and  $\hat{y}_{-i} = (n-1)\hat{x}/n$  is the accompanying

---

<sup>26</sup>Strausz (1999) shows for sequential partnerships in a model without norms that a sharing rule exists which uniquely implements efficient production. However, this scheme requires some liability of the agents as negative payoffs may occur.

<sup>27</sup>For a detailed analysis of noisy production in teams given a Leontief technology without social norms see Hvide (2001).

expected output from the others. This defines  $u_i$  as a random variable, since  $\bar{y}_{-i}$  is random. Hence, when each worker chooses her effort level, then this choice is made so as to maximize the expected value of  $u_i$ , conditional upon the others' (anticipated) equilibrium efforts.

Secondly, consider the case in which judgement is based upon consequences, not actions taken. A worker thus feels ashamed when his output is low even when he made a big effort. In this case we would obtain

$$u_i = w(y) - \frac{1}{2}x_i^2 - \frac{1}{2}(\hat{y} - y_i)^2 p(\hat{y}_{-i} - \bar{y}_{-i}), \quad (22)$$

where  $\hat{y}$  is the expected output of a single worker when exerting the ideal effort, and where  $u_i$  again is a random variable. This second case also covers situations in which the norm is externally enforced and workers observe each others' output (and care about others' opinions which, in turn, are based on their observations of individual outputs).

We believe that the qualitative features of the model remain intact under such generalizations at least when the stochastic shocks are moderate—for the simple reason that one obtains the deterministic model in the limit as the dispersion of the shocks is brought down towards zero.

**3.6. Heterogeneity.** We have so far assumed workers to be identical. An important question thus is how this analysis generalizes to the more realistic case of heterogeneous workers. In the present context, there seem to be essentially three dimensions in which heterogeneity is relevant: differing individual productivities, preferences for leisure, and sensitivity to the social norm, respectively.

All types of heterogeneity can of course be technically included in the above model. However, there are some subtle issues that have to be addressed. For example, if workers have different productivities, the socially efficient team optimal solution requires them to exert different degrees of effort. If this is captured by social preferences one might call them “utilitarian,” requiring from everyone his or her socially optimal effort. However, one could also imagine that workers feel that everybody should exert the same effort, in which case one might call the social preferences “egalitarian,” requiring from everyone the same socially optimal effort. Furthermore, the informational structure (who knows what about whom?) gains additional importance.

In the following we shall illustrate just one very simple case where workers only differ in their productivities, i.e., total output is now  $y = \sum_i \theta_i x_i$  where  $\theta_i$  is worker  $i$ 's productivity. We shall assume that the principal only knows the distribution of productivities while workers know each other's individual productivities. For a given linear contract  $(a, b)$  the socially optimal efforts are thus  $\hat{x}_i = \theta_i b$ .

We shall assume that social preferences are utilitarian. This implies that the subutility functions  $v$  can no longer simply depend on own effort and others' average effort, rather it must depend on the entire effort vector of all workers. Moreover, the issue arises whether deviations from socially optimal efforts are perceived in absolute or relative terms. We find it plausible that, given that social preferences are utilitarian and, hence, relative across workers, deviations are also perceived in relative terms. This means that we can define  $r_i = x_i/\theta_i$  as worker  $i$ 's relative effort and  $\bar{r}_{-i}$  as the average relative effort among the other workers. Worker  $i$ 's deviation from what the team optimum can now be measured by  $\hat{x}_i/\theta_i - x_i/\theta_i = b - r_i$ , and we may assume utility functions of the form

$$u_i(r_i, \bar{r}_{-i}, b) = a + \frac{b}{n} \sum_i \theta_i^2 r_i - \frac{1}{2} (\theta_i r_i)^2 - v(r_i, \bar{r}_{-i}, b),$$

where  $v$  is simply

$$v(r_i, \bar{r}_{-i}, b) = \frac{1}{2} (b - r_i)^2 p (b - \bar{r}_{-i}).$$

The first-order conditions become

$$r_i = \frac{np(b - \bar{r}_{-i}) + \theta_i^2}{p(b - \bar{r}_{-i}) + \theta_i^2} \cdot \frac{b}{n}.$$

If all workers productivities are equal to 1 we obtain, of course, our old fixed-point equation. Moreover, we see that the right-hand side is continuous in  $\theta_i$ . Hence, it is obvious that, at least for moderate productivity differences, all the results we obtained above for the homogenous case will carry over to the case where workers differ in productivity. Similar arguments can be applied to deal with heterogeneity in attitudes towards leisure and the norm itself.

**3.7. Trigger incentives.** In a deterministic setting, step-function contracts can be used to push efforts up such that any predetermined output target  $t$  is reached. For example, consider the step-function contract  $(s, t)$ , where

$$w(y) = \begin{cases} s & \text{if } y \geq t \\ 0 & \text{if } y < t \end{cases}. \quad (23)$$

In other words, each worker receives the pay  $s$  if the team produces at least  $t$ . If the pay  $s$  exceeds the required *per capita* effort,  $(t/n)^2/2$ , then all workers will find it in their best interest to exert exactly the required effort, given the linear-quadratic preferences we have assumed, and in the absence of a social norm. The owner can thus choose  $t$  so as to reach his or her first-best optimal profit.

However, we find such contracts less plausible than smoother contracts, such as the linear ones analyzed above. First, the knife-edge property of such a scheme makes it non-robust. If the team's output for some reason would fall just marginally below the target, perhaps due to some worker's "tremble" when exerting effort, or some small negative shock to the production technology, then all workers would stand there without any pay, although perhaps virtually all of them made the required effort. Secondly, there is a whole continuum of Nash equilibria for all  $s > (t/n)^2/2$ , and the slightest mis-coordination in this set may result in zero pay to all workers. Thirdly, such step-function contracts are vulnerable to collusion between the principal and a single worker. Suppose all workers plan to exert the minimal effort for reaching the target,  $x = t/n$ . A dishonest manager can then lure a worker to marginally reduce his or her effort, so that he or she, according to the contract does not have to pay anything to the team of workers, and secretly give a sizeable side payment to the colluding worker for this.<sup>28</sup>

In the absence of limited liability and wealth constraints, another contract form is available (also in a stochastic setting), by which the firm owner can implement the first-best solution. This can be done by essentially selling the firm to the workers, that is, by choosing a large *negative* "base salary" and a high bonus rate.<sup>29</sup> In our base model, this corresponds to a large negative  $a$  combined with the first-best bonus rate  $b = n$ . The parameter  $a$  then determines the rent distribution between the owner and the workers. As shown by Barron and Gjerde (1997), who analyze the impact of peer pressure in firms, given that such 100% marginal compensation rules are feasible, social norms may work against the interest of the owner.

#### 4. RELATIVE-PERFORMANCE PAY

In all the model versions considered so far, a worker's effort causes a positive externality for others—an increase in  $i$ 's effort increases  $j$ 's income, *ceteris paribus*. We saw that the efficiency-based group norm then works in favor of the firm owner. In other environments, such as when there is an element of competition between the workers, one worker's effort may cause a negative externality on other workers—an increase in  $i$ 's effort may decrease  $j$ 's income. Can a group norm then work against the owner's interest? If peer pressure makes workers compete less with each other, then the norm may restrain their efforts and cause profits to go down.<sup>30</sup>

In order to analyze this in the simplest possible setting, suppose that the owner observes each worker's effort and pays each worker a base wage and a bonus propor-

---

<sup>28</sup>A social norm among the workers, based on social efficiency for the team of workers, may work against such collusion, and thereby perhaps hinder it or make it less likely.

<sup>29</sup>This scheme was suggested by Groves (1973).

<sup>30</sup>We are grateful to Ernst Fehr who suggested this possibility.

tional to the worker's *relative* performance (granted at least one worker exerts some effort):

$$w_i = a + b \frac{x_i}{\sum_{j=1}^n x_j},$$

for some base wage  $a \geq 0$  and bonus rate  $b \geq 0$ .<sup>31</sup> (For the sake of definiteness, suppose the bonus payment is zero if all workers exert zero effort.) Note that an increase in a worker's effort reduces the bonus payment to the others. In this sense, each worker's effort imposes a *negative* externality on the other workers—the opposite of team incentives.

Assume the same linear-quadratic preferences over income and effort as in the basic model,<sup>32</sup>

$$u_i = a + b \frac{x_i}{\sum_{j=1}^n x_j} - \frac{1}{2} x_i^2, \quad (24)$$

and assume that the firm owner collects the residual,

$$\pi = y - b - na, \quad (25)$$

where  $y$  is the firm's output.<sup>33</sup>

It follows that the *team optimum effort level* now does not exist, strictly speaking: whenever at least one worker exerts positive effort, the team of workers earns  $na + b$ , so for the collective of workers it is better the less effort they make, as long as at least one worker makes some effort. However, when they all make exactly zero effort, then there is a discontinuous fall in their income - the team loses the whole bonus pay.

It is easy to verify that the unique symmetric Nash equilibrium - in the absence of a social norm and when efforts are chosen simultaneously - is that every worker exerts effort

$$x = \frac{1}{n} \sqrt{(n-1)b}. \quad (26)$$

As expected, this equilibrium effort is increasing in the bonus rate  $b$  and decreasing in the number  $n$  of workers.

---

<sup>31</sup>Such a payment scheme is also consistent with certain forms of asymmetric information, where the owner only observes a noisy signal of workers' efforts and  $w_i$  is the conditionally expected wage, given all workers' (true) efforts.

<sup>32</sup>Since income here is strictly concave in own effort, the disutility term could be taken as linear in effort.

<sup>33</sup>An interesting generalization of such relative-performance schemes is to let the bonus also depend on the firm's output an intermediate case between the present case and our basic case.

Next, let us include in the workers' preferences an efficiency norm in the same way as in the basic model:

$$u_i = a + b \frac{x_i}{\sum_{j=1}^n x_j} - \frac{1}{2} x_i^2 - \frac{1}{2} x_i^2 p(\bar{x}_{-i}), \quad (27)$$

where  $p : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is non-increasing: the less effort others do, the better for the team of workers, and the (weakly) more peer pressure does a worker feel, *ceteris paribus*. Note that in contrast to the case of team work, the social disutility felt by a worker now increases in his individual effort.

A necessary condition for an interior symmetric Nash equilibrium is thus

$$x = \frac{1}{n} \sqrt{\frac{(n-1)b}{1+p(x)}}. \quad (28)$$

We note that the right-hand side in this fixed-point equation is again non-decreasing in  $x$ , thus allowing for multiplicity of Nash equilibria, just as in the base model. Moreover, comparing (28) with (26), we see that the right-hand side is not larger (and typically smaller, namely if  $p > 0$ ) in the presence of social preferences of this form. Hence, as expected, equilibrium efforts are lower in the presence of a group efficiency norm. Somewhat speculatively, we conjecture that this observation might explain why relative performance pay is relatively rare in practice - the reason being that the social norm then works against the firm owner.<sup>34</sup>

What would be an optimal choice of tournament-contract  $(a, b)$  for the owner? It is clear from equation (28), that it is still optimal for the owner to set the base salary  $a$  equal to zero. However, the choice of bonus  $b$  will in part depend on the production technology. Suppose, for instance, that the production technology is the same as in the base model. Then, in any symmetric subgame equilibrium associated with a contract  $(0, b)$ , we have  $\pi = nx - b$ , where  $x$  is a Nash equilibrium effort level associated with the bonus rate  $b$ . In the absence of a social norm, this implies that the profit  $\pi$  is a strictly concave function of  $b$ , see equation (26), with maximum at  $b = (n-1)/4$ . This optimal contract - within the present class of linear relative-performance pay contracts - yields profits  $\pi = (n-1)/4$ .

**4.1. Example.** Assume that the peer pressure function is given by

$$p(x) = \frac{5}{1 + \exp[40(x - 0.2)]}.$$

---

<sup>34</sup>See Jensen and Murphy (1990) and Garen (1994).

Figure 3 shows the Nash equilibrium manifold (the S-shaped curve) for this logistic function with  $n = 5$ . Furthermore, the two straight lines represent iso-profit curves for  $\pi = 1$  and  $\pi = 0$ , respectively. Finally, the concave line represents the equilibrium correspondence in the absence of social preferences. The figure shows how social preferences press down workers effort levels, and also create multiple equilibria for certain bonus rates. With a social norm, there is a local profit maximum at  $b \approx 0.2$ , whereas the global profit maximum is at  $b \approx 1$ . Note that again a sufficiently high bonus rate (above  $b \approx 0.9$ ) knocks out the social incentive - this time to the benefit of the firm owner.

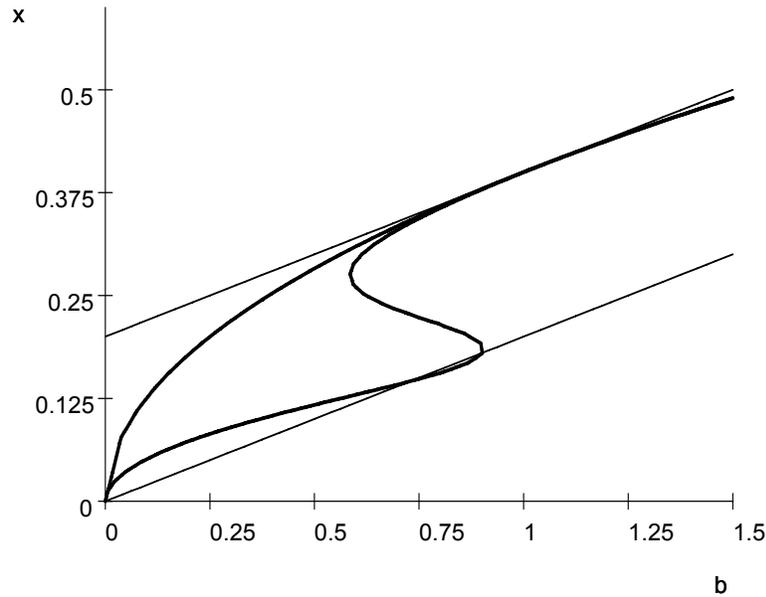


Figure 3: Nash equilibrium efforts  $x$  (vertical axis) for different bonus rates  $b$ , under relative-performance pay.

## 5. DISCUSSION

In this paper we first focus on incentive schemes for teams in the presence of a group norm. In our simple model, there is a moral hazard problem, due to the principal choosing a bonus rate which does not fully compensate the team of workers, and there is a free-rider problem among the team members. The work norm mitigates the free-rider problem by causing those whose effort falls short of the team optimum

to suffer a utility loss. We show that the equilibrium effort will never be lower than in the absence of a norm. Thus, a principal will benefit if he employs workers who are motivated by a norm.

We also find that multiple equilibria can arise. As a consequence there can be discrete jumps in effort levels even when the bonus rate is changed only marginally. This phenomenon provides an opportunity for dynamic wage setting, for example, in order to leave the basin of attraction of a low-effort equilibrium. Once a new equilibrium has been established, bonus rates may be restored to their old levels. Multiple equilibria can also cause rather nasty surprises for principals who raise the bonus rate. In some cases, a raise of the bonus rate destroys high-effort equilibria with the consequence that workers' efforts fall dramatically. This suggests a possible re-interpretation of so-called crowding-out effects, documented by Frey (1997) among others.

We also model the case of negative externalities between workers, which arise for example under relative-performance pay. With such incentive schemes, a group norm based on social efficiency for the workers demands workers to work less than what is individually optimal. While such a norm increases workers' utility, it reduces the firm's profit (and may hence endanger the firms' survival).

The analysis of the two different models, with wages based on team output and on relative performance, respectively, demonstrates how economic institutions may have direct and indirect effects on individuals' effort choices. In addition to the well-known direct effect by way of economic incentives, the institution may shape a social norm which, in turn, will also influence behavior. This suggests the importance of "norm management" by a principal who designs a contract.<sup>35</sup>

A key observation for understanding the evolution of work norms is that the matching between workers and firms is typically not random. Rather workers apply to selected firms and firms select applicants after careful interviewing. If ability was the only concern to the firms, interviewing would be simple and standardized. But this is not what we observe. On the contrary, firms care a lot about dimensions that can be summarized under "personality".<sup>36</sup> A version of our team production model with a social norm can explain why this is the case. As the equilibria in the effort game are Pareto-ranked, our firms would try to select workers who are sensitive to peer pressure. And firms that don't care for the "personality" of their workers would consistently earn less than others, and might therefore ultimately disappear. For workers, similar dynamics may apply. Those who feel little or no peer pressure

---

<sup>35</sup>See Kübler (2001) on the related notion of "norm regulation".

<sup>36</sup>See, for example, the recent article by Highhouse (2002) who discusses the advantages of the "holistic approach" to personnel selection over standardized tests.

would only be selected by firms with a lower "work morale", that is, by firms that in equilibrium pay less and that face a bigger risk of being shut down. This implies a double disadvantage for workers who free-ride. They earn lower wages and they are more likely to lose their jobs. Hence, there may be evolutionary selection in favor of workers who are sensitive to peer pressure in such settings. Interestingly, the opposite holds true for tournament production. A firm using relative performance schemes would like to select workers who are insensitive to social pressure. Thus, different incentive schemes can lead to a sorting of worker types, similar to effects observed between the private and the public sector.<sup>37</sup>

In this vein, Kandori (2002) highlights the dynamics of norms and morale in a model where both the norm itself and its strength evolve over time.<sup>38</sup> One of Kandori's conclusions is that workers should be allocated to branches of a firm so that those with similar levels of work morale are grouped together. This prevents decay of the work norm among those who are most socially oriented.

In a future project we plan to study team production game protocols in a series of laboratory experiments.<sup>39</sup> While we conjecture that this will confirm the relevance of social norms, we also expect that institutional detail will influence the evolution and strength of such norms.<sup>40</sup> Another avenue for future work is to use tools from evolutionary game theory to enable more powerful equilibrium selection (see Kandori (2002) for an analysis along those lines), and to go deeper into the modelling of social norm formation.

---

<sup>37</sup>See, for example, Francois (2000), who deals with public sector motivation from the vantage point of economics, or the recent survey on public sector motivation in the literature on public administration by Wright (2001).

<sup>38</sup>Compare footnote ??.

<sup>39</sup>In experimental labor markets, social norms based on reciprocity have been shown to be extremely important. See, in particular, the work by Fehr and collaborators, e.g., Fehr, Kirchsteiger, and Riedl (1993) or Fehr, Gächter, and Kirchsteiger (1997). However, these experiments deal with reciprocal relations (or norms) between an employer and an employee, while our study addresses social relations between employees.

<sup>40</sup>See, for example, recent work by Bohnet, Frey, and Huck (2001) who study the evolution of trust in contractual relationships, both theoretically and experimentally. They find that institutions matter directly (via short-run incentives) and indirectly (via changing agents' attitudes.) In an experiment by Andreoni (1995), contributions to a public good decrease significantly when the framing is changed from a setting with positive externalities to a setting with negative externalities, although the economic incentives remain unchanged. This asymmetry might have been caused by different norms that were triggered by the two respective frames.

## REFERENCES

- [1] Akerlof G. (1980). A Theory of Social Custom, of Which Unemployment May Be One Consequence, *Quarterly Journal of Economics* 84: 749-775.
- [2] Anderson, S.P., A. de Palma and J.-F. Thisse (1992). *Discrete choice Theory of Product Differentiation*, Cambridge and London: MIT Press.
- [3] Andreoni, J. (1995). Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. *Quarterly Journal of Economics* 110: 1-21.
- [4] Asheim, G. and M. Dufwenberg (2002). Deductive Reasoning in Extensive Games. *Economic Journal* (forthcoming).
- [5] Barron, J.M. and K.P. Gjerde (1997). Peer Pressure in an Agency Relationship. *Journal of Labour Economics* 15: 234-254.
- [6] Benaim, M. and J. Weibull (2002). Deterministic Approximation of Stochastic Evolution in Games, forthcoming, *Econometrica*.
- [7] Bernheim D. (1994). A Theory of Conformity. *Journal of Political Economy* 102(5): 841-877.
- [8] Besley T. and S. Coate (1992). Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination. *Journal of Public Economics* 48: 165-183.
- [9] Bohnet, I., B.S. Frey, and S. Huck (2001). More Order with Less Law: On Contract Enforcement, Trust, and Crowding. *American Political Science Review* 95: 131-144.
- [10] Encinosa, W.E., M. Gaynor and J.B. Rebitzer (1997). The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems. NBER Working Paper 5953.
- [11] Falk, A., U. Fischbacher and S. Gächter (2002). Isolating Social Interaction Effects. Mimeo. University of Zurich.
- [12] Falk, A., and A. Ichino (2003). Clean evidence on peer pressure. Mimeo.
- [13] Fehr, E., S. Gächter and G. Kirchsteiger (1997). Reciprocity as a Contract Enforcement Device. *Econometrica* 65: 833-860.

- [14] Fehr, E., G. Kirchsteiger and A. Riedl (1993). Does Fairness Prevent Market Clearing? *Quarterly Journal of Economics* 108: 437-460.
- [15] Fehr, E., K. Sigmund and M.A. Nowak (2002). The Economics of Fair Play. *Scientific American*.
- [16] Francois, P. (2000). 'Public Service Motivation' as an Argument for Government Provision. *Journal of Public Economics* 78, 275-299.
- [17] Frey, B.S. (1997). *Not Just for the Money: An Economic Theory of Personal Motivation*, Cheltenham: Edward Elgar.
- [18] Frey, B.S. (2003). Motivation Crowding Theory: A Survey of Empirical Evidence. *Journal of Economic Surveys* (forthcoming).
- [19] Garen, J.E. (1994). Executive Compensation and Principal-Agent Theory. *Journal of Political Economy* 102: 1175-1199.
- [20] Groves, T. (1973). Incentives in Teams. *Econometrica* 41: 617-631.
- [21] Hart, O. (2001). Norms and the Theory of the Firm. Harvard University. SSRN Electronic Paper Collection.
- [22] Highhouse, S. (2002). Assessing the Candidate as a Whole: A Historical and Critical Analysis of Individual Psychological Assessment for Personnel Decision Making. *Personnel Psychology* 55, 363-396.
- [23] Holmström, B. (1982). Moral Hazard in Teams. *Bell Journal of Economics* 13: 324-340.
- [24] Hvide, H.K. (2001). Some Comments on Free-Riding in Leontief Partnerships. *Economic Inquiry* 39: 467-473.
- [25] Ichino, A. and G. Maggi (2000). Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm. *Quarterly Journal of Economics* 115: 1057-1090.
- [26] Jensen, M.C. and K.J. Murphy (1990). Performance Pay and Top-Management Incentives. *Journal of Political Economy* 98: 225-264.
- [27] Kandel, E. and E.P. Lazear (1992). Peer Pressure and Partnerships. *Journal of Political Economy* 100: 801-817.

- [28] Kandori, M. (2002): The Erosion and Sustainability of Norms and Morale. The Japanese Economic Association-Nakahara Prize Lecture, mimeo.
- [29] Kandori, M., G. Mailath, and R. Rob (1993). Learning, Mutation, and Long-Run Equilibria in Games. *Econometrica* 61, 29-56.
- [30] Knez, M. and D. Simester (1999). Firm-wide Incentives and Mutual Monitoring at Continental Airlines. *Journal of Labor Economics* (forthcoming).
- [31] Kübler, D. (2001). On the Regulation of Social Norms. *Journal of Law, Economics, and Organization* 17: 449-476.
- [32] Legros, P. and S.A. Matthews (1993). Efficient and Nearly-Efficient Partnerships. *Review of Economic Studies* 68: 599-611.
- [33] Lindbeck, A., S. Nyberg and J. Weibull (1999): Social Norms and Economic Incentives in the Welfare State. *Quarterly Journal of Economics* CXIV: 1-35.
- [34] Lindbeck, A., S. Nyberg and J. Weibull (2002): Social Norms and Welfare State Dynamics. Forthcoming, *European Economic Review, Papers and Proceedings*.
- [35] Moffitt R. (1983). An Economic Model of Welfare Stigma. *American Economic Review* 73(5): 1023-35.
- [36] Rey Biel, P. (2002): Inequity Aversion and Team Incentives. Mimeo.
- [37] Strausz. R. (1999). Efficiency in Sequential Partnerships. *Journal of Economic Theory* 85: 140-156.
- [38] Titmuss, R.M. (1970). *The Gift Relationship*, London: Allen and Unwin.
- [39] Vislie, J. (1994). Efficiency and Equilibria in Complementary Teams. *Journal of Economic Behavior and Organization* 23: 83-91.
- [40] Weibull J. (1997). Norms in Firms: An Example. Mimeo. Stockholm School of Economics.
- [41] Wright, B.E. (2001). Public Sector Motivation: A Review of the Current Literature and a Revised Conceptual Model. *Journal of Public Administration Research and Theory* 11, 559-586.
- [42] Young, P. (1993). The Evolution of Conventions. *Econometrica* 61, 57-84.