Human Relations in the Workplace

Julio J. Rotemberg *

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

Human Relations in the Workplace

Julio J. Rotemberg *

Abstract

This paper seeks to understand what motivates workers to be friendly towards one another and studies whether firms benefit from encouraging these "human relations" in the workplace. The paper first proposes that feelings of friendship, or altruism, can be individually rational in certain settings where the variables controlled by the workers are strategically linked. The paper then studies what this implies for equilibrium altruism in three situations. The first has workers who are paid as a function of joint output. The second relationship I consider is that between subordinates and their supervisors while the third has workers paid as a function of their relative performance.

Human Relations in the Workplace

Julio J. Rotemberg

People have feelings for those they work with. This raises two questions. The first is whether these feelings affect performance on the job and the second, whose interest is enhanced if the answer to the first question is affirmative, is what gives rise to these feelings. In this paper I consider both the effects and the causes of one particular set of feelings, namely feelings of friendship and altruism. Having answered both questions, I give conditions under which a firm would increase its profits by spending resources to enhance the friendliness of employees towards each other.

The idea that feelings of friendship (or "human relations") affect performance in the workplace is an old one, dating back at least to Mayo (1933). The famous "Hawthorne experiments" reported of Roethlisberger and Dickson (1939) appeared to confirm the importance of these personal relations in industrial settings. Subsequently, a large number of empirical studies were conducted relating productivity in small groups to the feelings reported by the members of the group. One particular feature of groups that has been explored in several studies (see Goodman, Ravlin and Schminke (1985) for a discussion) is group cohesiveness, whose definition varies but which often means the extent to which members of the group express that they like other members of the group. Unfortunately, these studies of the relation between cohesiveness and productivity have delivered mixed results.

In this paper, I equate people's expression of liking for each other (*i.e.*, those which lead groups to be classified as highly cohesive) with feelings of altruism. The ambiguity of the relationship between cohesiveness and productivity is then not surprising. If employees who like each other take actions who benefit fellow employees, (*i.e.*, act altruistically), the effects on firm output and firm profitability depend on the setting. Altruism towards fellow employees may lead an employee to work little or to work hard. It has the former effect if a cutback in an individual's effort allows fellow employees to reduce their own effort while keeping their income constant. It has the latter effect if, by working harder an employee raises the income of his or her fellow employees. This

1

would occur if individual output is unobservable and each employee is paid as a function of group output.

This ambiguity is the same as that analyzed by Holmström and Milgrom (1990) in their study of "collusion" in organizations. They focused on the effects on firm profitability of allowing workers to collude *i.e.*, to write binding contracts with each other. Like altruism, these binding contracts also lead workers to take actions that enhance each other's welfare. Holmström and Milgrom show that, when the firm's best way of inducing effort by employees is to make them to compete with each other, as in Lazear and Rosen (1971), collusion is bad for the firm because it leads both employees to exert lower effort. On the other hand, if the firm can only observe collective output so that compensation is tied to group performance, collusion is beneficial to the firm.

One issue that arises at this point is why one should study altruism in the workplace if its effects are similar to those of contracts. One reason is that contracts between workers, particularly if they hurt the employer are not enforceable by the courts. The contracts might be self enforcing as a result of the repeated nature of the interaction. One key difference between this self-enforcement and altruism is that altruism works even when the relationship between workers is about to end. Another difference is that, as I will argue below, the conditions under which altruism is likely to arise are not the same as those under which workers profit from contracting with each other. Finally, my model of altruism generally has a unique equilibrium whereas equilibria in self-enforcing repeated interactions are typically non-unique.

A second issue is whether the mixed results in the empirical relationship between the cohesiveness of groups and their productivity can be explained by equating cohesiveness with a high degree of altruism. In other words, do people who express that they like their co-workers behave altruistically towards them? I argue below that some of these empirical studies are indeed consistent with the view that cohesion has the same effect on output as altruism. However, there are many other empirical studies of this relationship which do not provide sufficient details to verify whether they are consistent with the model developed here.

If mutual attraction, and altruism, are indeed relevant in affecting performance within the workplace, it becomes important to ask what brings these feelings about. The existing theories of attraction between individuals are generally "backwards looking". They treat the liking of an

2

individual for another as a reaction to what has gone on before. Thus, there exist large literatures (see Aronson (1984) for references) showing that, in experimental settings, people tend to like people whom they perceive to be similar to themselves and people who are physically attractive. Along similar lines, Homans (1950) and Thibaut and Kelley (1959) posit that attraction is a natural consequence of having pleasant interactions with people. A general conclusion that seems to emerge from these observations is that people like individuals who give them large "rewards" at low "costs" (Homans (1961). Thus physically attractive people confer aesthetic rewards, people that agree with us confer the reward of making us feel that we are right etc.

This general conclusion raises the question whether people benefit by liking those who give them large rewards. In other words, are feelings of affection instrumental in raising one's own utility? In this paper I pursue this possibility; I develop a model in which people become altruistic when this is in their self interest. Formally, I treat the degree of altruism as a choice variable: an individual starts with a utility defined over his own welfare and chooses whether to change his utility function and become altruistic.[1] If he does change it, his decisions from then on take into account the welfare of another individual.

The assumption that people choose their altruism, which may be controversial, can be thought of in two other ways. First, one can think of the individual as taking actions (such as inviting the other to dinner) which modify his attitude towards the other person. Viewed in this way, the model is akin to the rational addiction model of Becker and Murphy (1988) where the individual knows that the initial consumption of an addictive substance will change his attitude towards the substance in the future. Second, one can think of natural selection as favoring individuals who change their emotions whenever that is useful to them. In this case, the individual might feel no control over his emotions even though these are changing in self interested ways.

I now turn to the benefits that the individual derives from his altruism. It should be clear that, unless this altruism can be credibly demonstrated, it is not in the individual's self interest. Altruism leads the individual to take actions which maximize a combination of his own and the

---

[1] The sort of opportunistic change in tastes I consider is also featured in Akerlof (1983) who shows how people benefit from becoming loyal, in Frank (1987) who shows that they benefit from having feelings of guilt when they take advantage of others. Coleman's (1990) self-interested change in tastes is somewhat different since it involves an actor who becomes happy by "changing himself to be satisfied with the world" (p.517). Coleman (1990) also considers actors who relinquish control of their actions to groups of individuals. This is more similar to what is considered here since these groups have potentially different objectives from those of the narrowly conceived individual.

other's utility function and this results in a lower level of own utility than if the individual had been purely selfish.

On the other hand, consider the case where A can prove to B that he feel altruistic towards him. Knowing this, B knows that A's behavior will be different. This, in turn, will sometimes affect B's behavior. If B is led to modify his behavior in a way that benefits A, the initial decision by A to feel altruistic towards B will have been a smart one. I will show that the condition under which this occurs is related to the strategic complementary of Bulow, Geanakoplos and Klemperer (1985). This condition requires that, if each individual takes one action and these are normalized so that increases in one person's action raise the other person's welfare, a marginal increase in A's action leads to an increase in B's.

Since altruism is only beneficial if it can be proved, it must be irreversible, at least to some degree.[2] This irreversibility seems natural if one thinks of altruism as a change in tastes. Once a person is altruistic towards another, he stops having an incentive to alter the utility towards selfishness (unless that somehow made both better off). In practice, one does see people who stop liking one another. These further changes in emotions seem to arise as conditions within the relationship change, usually in ex ante unpredictable ways. To keep the analysis simple, my model has individuals interacting only once so that these later changes in emotions play no role.

This basic idea that altruism can be privately beneficial is closely related to Becker (1974a). Becker's (1974a) "rotten kid theorem" establishes that the interaction between a selfish and an altruistic individual can lead to Pareto Optimal allocations. In particular, the selfish individual will, under certain circumstances, take actions which maximize joint income. This result hinges on three assumptions. First, the altruistic individual must be so rich and so altruistic that, in equilibrium, he makes positive transfers to the selfish one. Second, the altruistic individual must act after the selfish one. These assumptions imply that the altruist will lower his transfers to the selfish individual when the latter reduces joint income.[3] Finally, the selfish individual's utility must be linear in the ex post transfers by the altruistic one. (Bergstrom (1989)).

The "rotten kid theorem" does not establish that the altruist gains from his altruism. In a closely related paper, Becker (1976) argues that altruistic individuals can, indeed, end up better

---

[2] It is thus a form of commitment.
[3] Further discussion of these two assumptions can be found in Hirshleifer (1977).

off than selfish ones. His argument is largely informal, though it too is based on an example where the altruist would, ex post, punish the selfish individual if the latter failed to take actions which benefit the altruist. On the other hand, Bernheim and Stark (1988) provide an example where selfish individuals do better than altruists.

Given these contradictory findings, I devote Sections 1 and 5 to clarifying when altruism is individually rational.[4] To avoid some of the asymmetries introduced when players take actions sequentially, I focus mostly on games with simultaneous moves. Only in Section 5 do I discuss sequential moves. With this change in the timing of moves, strategic complementarity ceases to be sufficient for equilibrium altruism, although it remains a force which promotes such altruism.

After considering rational altruism within abstract simultaneous move games in Section 1, I turn to its applications in the workplace. This boils down to an analysis of whether the strategic variables of different employees tend to be strategic complements or substitutes in different settings. When they are complements, friendship should arise endogenously.

The first relationship I consider is that between two employees who work as a team in that their payment depends only on their joint output. I show that, as a result, altruism tends to arise and raise productivity. Thus the issue whether firms benefit by spending resources to increase the friendliness of their employees is a subtle one. In the case of joint output, where the firm benefits from altruism among employees, it arises anyway as a consequence of the dependence of employee compensation on total output. Moreover, the firm can make sure the employees like each other by changing the group's incentive payment. Nonetheless, I show that there are circumstances where additional payments to foster altruism are beneficial to the firm.

Section 2 also contains an extended discussion of some of the "Hawthorne experiments" reported by Roethlisberger and Dickson (1939). These experiments suggest that changing the incentive payment together with the creation of an atmosphere conducive to friendship help productivity more than either change on its own. I also discuss additional evidence on the relationship between cohesiveness and productivity which shows that altruism and measures of group cohesiveness are closely linked.

In Section 3, I consider a second relationship within the workplace and that is the relationship

---

[4] Altruism is rational in my model in the same sense that addiction is rational in Becker and Murphy (1988).

of authority. Several authors including Homans (1950) have pointed out that relationships of authority tend to be less friendly than relations among more equal co-workers. I show that, in the case of supervisors whose main role is to monitor their subordinates' actions for purposes of compensation, such absence of friendship is to be expected. I also consider an authority relationship described in Crozier (1964) in which a "team leader" has no authority to set compensation but does have some control over the pace of work. I focus on this relationship because it has an unexpected feature. The workers in this setting generally reported that they did not like each other (which I equate with lack of altruism). The exception was that team leaders liked their co-workers. This assymetry seems hard to explain with the backwards-looking models of altruism but fits well with my own model.

Third, in Section 4, I consider two workers whose individual output is observable. As in Lazear and Rosen (1971) there is a common shock to productivity which is not observable by the firm. Thus, payment schemes that pay workers as a function of the rank order of their output, are quite desirable. Yet, Lazear (1989) shows that they are relatively rare. I show in Section 4 that these schemes often encourage altruism between the affected workers and thereby lose some of their desirable incentive effects. Section 5 considers sequential games while Section 6 offers some conclusions.

1. Rational Altruism

In this section, I show how individuals can benefit by changing their tastes towards altruism. I take up separately the general case of continuous choices and the classical prisoner's dilemma problem.

1.1. Continuous Choices

There are two individuals, A and B, who carry out actions a and b respectively. The initial utility functions of A and B are, respectively, $\alpha(a,b)$ and $\beta(a,b)$. These utility functions can be thought of as being the "true" welfare (or, in the terminology of Rabin (1991), "material payoff") of A and B so that α depends only on what happens to A individually and similarly for β. Thus, in my simple settings α depends on A's income and on his own effort but not on B's income or effort.[5] More generally, particularly if one takes an evolutionary interpretation of changing tastes,

---

[5] Assuming that the disutility of effort is separable from whatever pleasure the individual gets as a result of his altruism, the

6

one could define α and β as the ultimate fitness for survival of the two agents.

If the utility functions remain α and β, the Nash equilibrium pair of actions a^n, b^n solve the following standard equations

$$\alpha_a(a^n, b^n) = 0 \tag{1.1}$$

$$\beta_b(a^n, b^n) = 0 \tag{1.2}$$

where subscripts denote partial derivatives. The Nash equilibrium is generally not Pareto Optimal. Pareto optimality requires that, for some positive λ the pair of actions (a, b) satisfy

$$\alpha_a(a, b) + \lambda\beta_a(a, b) = 0$$
$$\alpha_b(a, b) + \lambda\beta_b(a, b) = 0 \tag{1.3}$$

Thus, Pareto optimality holds at the Nash equilibrium only if $\alpha_b(a^n, b^n) = 0$ and $\beta_a(a^n, b^n) = 0$.

I assume that, before A chooses a, he chooses a taste parameter γ_A while B chooses γ_B. After choosing these parameters, the choices of a and b maximize U_A and U_b respectively where these *ex post* utility functions are given by

$$U_A = \alpha(a, b) + \gamma_A U_B \tag{1.4}$$

$$U_B = \beta(a, b) + \gamma_B U_A. \tag{1.5}$$

Thus, the parameters γ_A and γ_B represent the degree of altruism chosen by A and B to maximize α and β respectively. Instead of having people choose parameters, one could obtain equivalent results by having them choose actions such that, as in Becker and Murphy (1988) the preferences are changed by the actions.

Equations (1.4) and (1.5) can be solved for U_A and U_B yielding

$$U_A = \frac{\alpha + \gamma_A\beta}{1 - \gamma_A\gamma_B}$$
$$U_A = \frac{\beta + \gamma_B\alpha}{1 - \gamma_A\gamma_B} \tag{1.6}$$

It is apparent from (1.6) that this problem is not well defined unless γ_A and γ_B are less than one. Before maximizing α and β with respect to the γ's I develop intuition for the effects of changes

---

measurement of α and β is relatively straightforward.

in tastes by analyzing the effect of changing γ_B on both α and β. This is obviously different from asking how U_A or U_B respond to changes in γ_B. This raises the question of whether one should assume that the γ's are picked to maximize the U's instead of α and β. The former is closer to Becker and Murphy (1988) and Montgomery (1991) since they assume that agents at t evaluate future outcomes with the utility function that is modified by the actions taken at t instead of evaluating them using the period t utility function itself. I pursue the latter approach for two reasons.

First, having people choose γ to maximize U implicitly treats A as choosing whether to be friends with B or not to have any friends at all. A more plausible view is that A is sure to have some number of friends and that he must decide whether B (or someone else) should be in this set. Thus, one can imagine that A's total benefits from all his friendships are independent of γ_A even if one equates $\gamma_A U_B$ with the psychological benefits to A from his friendship with B.[6] This would occur if a higher γ_A reduces the utility from friendship with other people. In this case, A ought to pick γ_A to maximize α even though his *ex ante* evaluation of outcomes is the same as his *ex post* evaluation.

Second, and perhaps related to this weakness, having people choose their γ's to maximize the U's has several unpleasant implications. As Bernheim and Stark (1988) show, changes in the γ's can have paradoxical effects on the U's because of their effect on the denominators of (1.6). Also, people who are free to choose their γ would set it equal to $+\infty$ if their partner is happy and equal to $-\infty$ is he is unhappy.[7]

Given the preferences in (1.4) and (1.5), the Nash equilibrium actions satisfy

$$\alpha_a + \gamma_A \beta_a = 0 \tag{1.7}$$

$$\beta_b + \gamma_B \alpha_b = 0 \tag{1.8}$$

Comparing (1.3), (1.7) and (1.8), it is apparent that, in the limit where γ_A and γ_B equal one, the Nash equilibrium with altruism is Pareto Optimal. Supposing in addition that the solution to (1.7) and (1.8) when $\gamma_A = \gamma_B$ is monotone in the common value of γ, it follows that increasing γ

---

[6] Using Rabin's (1991) terminology these would be the "psychological payoffs" from the relationship. The benefits from friendship are probably more complex than an expression like $\gamma_A U_B$. I use this formulation as a starting point because it is standard in the literature on altruism.

[7] Montgomery (1991) avoids this problem by postulating that the γ's depend on previous choices of consumption in a way that makes infinite values either impossible or very undesirable to the individual.

leads to actions that are closer to the Pareto Optimum. This explains why simultaneous increases in altruism are beneficial to the individuals. However, I am more concerned with whether an individual would unilaterally decide to become altruistic.

To study this, I now analyze the effect of changes in γ_B. Consider first the case where A is unaware of B's change of heart. Then, a remains constant and the change in b can be obtained by differentiating (1.8)

$$(\beta_{bb} + \gamma_B \alpha_{bb})db = -\alpha_b d\gamma_B \qquad (1.9)$$

The second order conditions for the maximization of utility ensure that the expression multiplying db is negative. As long as α_b does not change sign as one varies γ_B, this implies that increases in γ_B hurt B. The reason is that they move b further and further away from the point that maximizes β.

Now consider the effect on A's ex ante utility. This is given by

$$\alpha_b db = -\frac{(\alpha_b)^2}{\beta_{bb} + \gamma_B \alpha_{bb}} d\gamma_B > 0 \qquad (1.10)$$

so that A is always better off when B cares more for him.

I now turn to the more plausible case where B can demonstrate his affection. Given that B loses when his affection is known only to him, any gains to B from his affection must be due to the changes this affection induces in A's behavior. But, A will change his actions only if B can demonstrate his affection credibly, so that A is sure that B's actions will be affected. Such credibility requires that it be difficult (if not impossible) for a person in B's position who does not feel altruistic towards A to mimic the signal emitted by a true altruist.

And what signals should A believe? There are at least two types of somewhat credible signals. The first type, body language, is stressed by Homans (1950) who says : "In deciding what sentiments a person is feeling, we take notice of slight, evanescent tones of voice, expressions of his face, movements of his hands, ways of carrying his body, and we notice these things as part of a whole... From these wholes we infer the existence of internal states of the human body and call them anger, irritation, sympathy... [Yet] we act on our inferences, on our diagnoses of the sentiments of other people, and we do not always act ineffectively" (p. 39). Another way B can signal his affection is through favors and presents. This idea is pursued in Camerer (1988) who gives several reason why

gifts can be credible signals of affection. One of these reasons is that it may be more difficult for a person that does not actually care for another to deliver equally touching favors and gifts.

For simplicity, I neglect the costs of signaling altruism in most of my discussion. With costless signaling, the effect of a change in γ_B on both actions can be obtained by differentiating (1.7) and (1.8).

$$D \begin{pmatrix} da \\ db \end{pmatrix} = \begin{pmatrix} 0 \\ -\alpha_b \end{pmatrix} d\gamma_B \tag{1.11}$$

where

$$D = \begin{pmatrix} \alpha_{aa} + \gamma_A \beta_{aa} & \alpha_{ab} + \gamma_A \beta_{ab} \\ \beta_{ab} + \gamma_B \alpha_{ab} & \beta_{bb} + \gamma_B \alpha_{bb} \end{pmatrix}$$

Thus, the effect on a and b is given by

$$da = \alpha_b [\alpha_{ab} + \gamma_A \beta_{ab}]/|D| \tag{1.12}$$

$$db = -\alpha_b [\alpha_{aa} + \gamma_A \beta_{aa}]/|D| \tag{1.13}$$

and, using the equilibrium conditions (1.7) and (1.8) to substitute for α_a and β_b, the effects on α and β are given by

$$d\alpha = \alpha_a da + \alpha_b db = \frac{-\gamma_A \beta_a \alpha_b [\alpha_{ab} + \gamma_A \beta_{ab}] - \alpha_b^2 [\alpha_{aa} + \gamma_A \beta_{aa}]}{|D|} \tag{1.14}$$

$$d\beta = \beta_a da + \beta_b db = \frac{\alpha_b \beta_a [\alpha_{ab} + \gamma_A \beta_{ab}] + \gamma_B \alpha_b^2 [\alpha_{aa} + \gamma_A \beta_{aa}]}{|D|} \tag{1.15}$$

To obtain a determinate sign for $|D|$, I suppose that the solution is stable under the usual tatônnement process where da/dt depends positively on dU_A/da while db/dt depends positively on dU/db. This stability, which can be viewed as implying that people converge to the Nash equilibrium when they act in a fairly myopic manner, implies that the matrix D is negative semidefinite.[8] So, as long as D is not singular, its determinant is positive. This means that, starting at the point where γ_A and γ_B both equal zero (so that A and B are selfish) the sign of $d\beta$ is equal to the sign of $\alpha_b \beta_a \alpha_{ab}$. By appropriately normalizing the actions, we can ensure that α_b and β_a both have the same sign. In other words, we set up the problem so that increases in both actions are good for

---

[8] The reason is that, locally, the system of differential equations governing a and b is then given by

$$\begin{pmatrix} da/dt \\ db/dt \end{pmatrix} = D \begin{pmatrix} a - a^* \\ b - b^* \end{pmatrix}$$

where a^* and b^* are the equilibrium values.

the other player. Then, the sign $d\beta$ is the sign of α_{ab}. This second derivative is positive if a selfish A would want to increase a if he knew that B had increased b. Using Bulow, Geanakoplos and Klemperer's (1985) terminology, a positive α_{ab} makes a and b "strategic complements" from A's point of view. This strategic complementarity then leads B to become altruistic toward A.

The intuition for this result is straightforward. If B starts to feel altruistic towards A, he will move b in the direction a likes. Suppose we normalize the actions so that this direction is also the direction in which A must move his action to benefit B.[9] Then, if the actions are strategic complements, A does indeed move his action in this direction and B benefits from liking A. If, instead, A views the actions as strategic substitutes, B loses when he lets A know that he cares for him because this leads A to adjust a in the direction that B finds deleterious.[10]

I now turn to the effects of B's increased altruism on A's welfare. Perhaps surprisingly, these effects are ambiguous, although clear effects do arise when the altruism is purely one-sided so that A does not care for B. In this case, γ_A is zero so $d\alpha$ has the same sign as $-\alpha_{aa}$ which is positive. A is thus better off as B likes him more. However, if γ_A is positive, and a and b are strategic complements, the first term in (1.14) is negative. This is easiest to understand if one imagines that α_b and β_a are both positive and that the two actions are strategic complements for both individuals. In this case, a and b both increase as γ_b goes up. Once A likes B a great deal, further increases in γ_B can be deleterious to A's welfare because they induce excessive increases in a. This requires that γ_A itself be large. So the question becomes whether it is plausible that γ_A is ever so large that increases in γ_B hurt A.

I attempt to answer this question by assuming that each individual picks his own level of altruism to maximize his own ex ante level of utility. Thus γ_B is chosen to maximize β while γ_A maximized α. I then ask whether a small increase in γ_B starting from this point hurts A. To further simplify this analysis, I assume that the interaction between A and B is symmetric so that, for all a and b, $\alpha(a, b) = \beta(b, a)$. Then, at a symmetric equilibrium, the γ's are set so that (1.15)

---

[9] In Bulow, Geanakoplos and Klemperer (1985)'s work on strategic interactions between firms, there is no natural analogue to this normalization. For this reason, their results depend not only on whether actions are strategic complements or not. Instead, they depend also on whether prices or quantities are the strategic variable, whether there are economies or diseconomies across markets etc.

[10] Bernheim and Stark (1988) provide an example where increased altruism by B makes B worse off for this reason. Their example is based on the "Samaritan's Dilemma", B's altruism leads A to expect larger transfers in the future. This expectation in turn induces A to consume more in the present period (which is inefficient).

11

equals zero. Letting $\gamma = \gamma_A = \gamma_B$, we have

$$(\alpha_{aa} + \gamma\alpha_{bb})\gamma + \alpha_{ab}(1 + \gamma) = 0 \tag{1.16}$$

where the first term is negative while the second term is positive in the case of interest (where the variables are strategic complements). On the other hand, the sign of $d\alpha/d\gamma_B$ is the opposite as the sign of

$$(\alpha_{aa} + \gamma\alpha_{bb}) + \alpha_{ab}\gamma \tag{1.17}$$

Comparing (1.16) and (1.17) it is apparent that, as long as γ is smaller than one, the negative term is bigger in (1.17) than in (1.16) while the positive term is strictly smaller. Therefore, (1.17) is negative. This implies that $d\alpha/d\gamma_B$ is positive at the symmetric equilibrium. This result makes intuitive sense; altruism is something that tends to benefit others as well as oneself. So, one expects there to be too little altruism in equilibrium, not too much.

1.2. The Discrete Prisoner's Dilemma

In this subsection I consider the standard model of conflict, namely the Prisoner's dilemma. The payoffs in this game are given by

$$
\begin{array}{ccc}
 & C & D \\
C & 1 \quad , \ 1 & -\ell, \ 1+g \\
D & 1+g, \ -\ell & 0 \quad , \ 0
\end{array}
\tag{1.18}
$$

where ℓ and g are strictly positive.[11] I will treat the row player as A and the column player as B.

Since each player's optimal action (D) does not depend on the other player's action, there appear to exist no strategic complementarities in this game. If B starts liking A and this leads him to play C, he is strictly worse off if his altruism is not reciprocated. B ends up with $-\ell$ instead of zero. But, endogenous altruism can still affect the outcomes. Suppose that A and B have utilities given by (1.4) and (1.5). Then, using (1.18) to obtain the values of α and β, their perceived payoffs U become proportional to

$$
\begin{array}{ccc}
 & C & D \\
C & 1+\gamma_A \quad , \ 1+\gamma_B & -\ell + \gamma_A(1+g), \ 1+g-\gamma_B\ell \\
D & 1+g-\gamma_A\ell, \ -\ell+\gamma_B(1+g) & 0 \qquad , \ 0
\end{array}
$$

---

[11] The zeros and ones are simply normalizations. In any symmetric Prisoner's Dilemma, one can always subtract the same amount from each entry to make the $\{D, D\}$ entry equal to $\{0, 0\}$ and then multiply all the entries by a positive constant to make the $\{C, C\}$ entry equal to $\{1, 1\}$. The resulting game is strategically equivalent.

Consider first the symmetric situation where $\gamma_A = \gamma_B = \bar{\gamma}$. Then $\{C, C\}$ is a Nash equilibrium as long as

$$1 + \bar{\gamma} \geq 1 + g - \bar{\gamma}\ell \tag{1.19}$$

Since $\bar{\gamma}$ must be below one, this is possible only if $g/(1 + \ell)$ is less than one.

Playing C becomes a dominant strategy (so that $\{C, C\}$ is the only Nash equilibrium) if, in addition

$$\bar{\gamma}(1 + g) - \ell > 0 \tag{1.20}$$

which is possible with $\bar{\gamma} < 1$ only if $\ell/(1 + g)$ is also smaller than one.

Suppose that condition (1.19) is satisfied while (1.20) is violated. This implies that

$$\frac{g}{1 + \ell} < \bar{\gamma} < \frac{\ell}{1 + g} \tag{1.21}$$

Whether there exists a $\bar{\gamma}$ that satisfies (1.21) is crucial in what follows. It is thus worth noting that such a $\bar{\gamma}$ exists if $\ell > g$ and not otherwise.

If $\bar{\gamma}$ satisfies (1.21), both $\{C, C\}$ and $\{D, D\}$ are Nash equilibria. However, both equilibria are not equally plausible since $\{C, C\}$ is Pareto dominant. So, $\{D, D\}$ is not robust in the sense of Farrell and Saloner (1985). The reason is that if A and B make their choices sequentially, rather than simultaneously, the only subgame perfect equilibrium is $\{C, C\}$. If, say, A chooses first, he knows that he guarantees himself the superior outcome $\{C, C\}$ by choosing C.

Now suppose that B sets γ_B slightly above $g/(1 + \ell) < 1$ while A sets γ_A to zero so that affection goes only from B to A. If $\ell < g$, (1.21) fails for $\gamma_B = \bar{\gamma}$ and B plays C for sure. Since A is selfish and plays D, B's affection cost him ℓ; he could have gotten 0 by remaining selfish himself. If, instead, $\ell > g$, (1.21) is satisfied for this γ_B and the only Nash equilibrium is $\{D, D\}$ so that B's altruism is costless. Somewhat more generally, if a $\bar{\gamma}$ exists such that (1.21) holds, B can set $\gamma_B = \bar{\gamma}$ at no cost to himself even if A is not altruistic. If no such $\bar{\gamma}$ exists, B's affection is generally costly to him unless A reciprocates.

Now consider the initial stage where each individual picks his level of altruism. If $g < \ell$ and $\frac{g}{1+\ell} < 1$ then (1.21) can be satisfied for some $\bar{\gamma}$ less than one and choosing such a γ is a dominant strategy for both players. This choice gives a player the outcome $\{C, C\}$ if the other player also chooses a γ in this region while it costs nothing if the other player chooses a lower γ. On the other

13

hand, if $g > \ell$, the only equilibrium has the players remaining selfish. This is now a dominant strategy. By doing so one gains $1 + g$ if the other player becomes altruistic, while one loses nothing if the other player remains selfish.

The result is intuitively appealing since it says that people will tend to like each other when not doing so will hurt the other more than one would gain oneself by remaining selfish. What causes this result is that, if $\ell > g$, there are strategic complementarities once one player feels altruistic towards the other. In particular, if B likes A and $\ell > g$, then B responds to C with C while he responds to D with D. This strategic complementarity can in turn be traced to the fact that A's benefit from playing D rather than C does depend on whether B plays C or not. If B plays C, A's gain from playing D is g whereas it is ℓ if B plays D. So when ℓ exceeds g, A has a lower incentive to play D when B plays C.

So far, I have considered abstract games with payoffs given either by the standard Prisoner's Dilemma or by arbitrary continuous functions α and β. In the next sections, I study specific features of the payoffs α and β that arise inside firms. These payoffs are made to depend on the relationship of the two people within the organization. Before developing this theme, it is important to stress that personality also plays a role in the determination of the equilibrium values of α and β. In particular, the private costs of the actions a and b are likely to depend on the individual. For instance, an individual who is not very interested in career advancement (perhaps because he finds his leisure very valuable) would have a different attitude towards changing his effort than one who is very keen on professional progress. Thus, one should not expect the pattern of friendship and altruism to depend solely on the nature of the interactions in the workplace. With this caveat, I keep personality factors constant and analyze the effect of these interactions on sentiments.

2. Joint Production

Often, it is not possible to disentangle the effort two employees make in producing a good. As Holmström (1982) shows, if they are paid as a function of total output alone and they are selfish, their effort is subobtimally low. The reason is that each employee neglects the positive effects on the other employee's compensation of an increase in his own effort. I will consider two models of this phenomenon. The first subsection deals with the case of continuous actions and a differentiable production function relating output to these actions. The second subsection considers instead a

14

case where output depends only on two discrete levels of effort. I show that in both cases altruism emerges endogenously though the social optimum can only be achieved in the second setting. The third and fourth subsections consider experimental evidence on cohesiveness and productivity and relates it to the models of the first two subsections.

2.1. A Differentiable Production Function

Suppose output is given by the symmetric, monotone and increasing function $f(a, b)$ where a and b, once again represent the two actions. If, together, the two employees receive a price of one for their output and they split evenly the resulting revenue, the individual payoffs are

$$\alpha(a, b) = f(a, b)/2 - e(a)$$
$$\beta(a, b) = f(a, b)/2 - e(b) \tag{2.1}$$

where the e functions capture the cost of effort and with $e' > 0$, $e'' > 0$ where primes denote derivatives. The selfish Nash equilibrium then has $f_a = f_b = 2e'$ whereas the optimum requires that f_a and f_b be equal to e' itself. Since the marginal product of effort is smaller at the optimum, effort is higher there than at the selfish Nash equilibrium.

If the two employees act as if they had preferences given by (1.4) and (1.5), the equilibrium effort satisfies instead

$$f_a = \frac{2}{1 + \gamma_A} e'$$
$$f_b = \frac{2}{1 + \gamma_B} e'$$

so that the marginal product of each type of effort declines as the γ's rise. In the limit where the two γ's equal one, we obtain the same outcome as at the optimum.

Since α_{ab} equals f_{ab}, we can be sure that there is some altruism in equilibrium if, as in the Cobb-Douglas case, f_{ab} is positive. I now consider the symmetric equilibrium where each person chooses his γ. This equilibrium must satisfy (1.16). Thus

$$(f_{aa} - 2e'' + \gamma f_{bb})\gamma + f_{ab}(1 + \gamma) = 0 \tag{2.2}$$

As long as f is concave, the expression on the left hand side of (2.2) is negative for γ equal to one. We know this because concavity implies that $f_{aa} + f_{bb} + 2f_{ab}$ is negative while $(-e'')$ is

15

negative as well. On the other hand, if f_{ab} is positive, the expression on the left hand side of (2.2) is positive for γ equal to zero. By continuity, the solution to (2.2) therefore involves a γ strictly between zero and one. There is some altruism in equilibrium but not enough to reach the first best.

Because equilibrium altruism raises output and traditionally measured productivity, it tends to be in the firm's best interest as well. The firm benefits if the price it pays the team per unit of output (which I normalized to one) is less than the value to the firm of an addition to output. As long as the production of output involves some proprietary knowledge or even the use of some of the firm's capital, this will generally be true so that the firm is better off if the employees like each other than at the selfish Nash equilibrium. In this subsection, there is no mechanism for the firm to promote this cohesiveness nor have I looked at the tradeoff between promoting cohesiveness and raising the price paid employees per unit of output. I consider this in the next subsection within the context of a somewhat simpler model.

2.2. Discrete Levels of Effort and the Firm's Role in Promoting Cohesiveness

In this subsection I give an example where the firm is better off promoting cohesion, even at a cost, than changing its incentives for effort. To demonstrate this, I determine endogenously the payment per unit of output which was exogenously fixed at 1 in the previous subsection. In addition, I will incorporate explicitly the costs of signaling one's affective state which I have neglected so far. For simplicity, I now assume that the actions a and b can take only two values. Either the individual makes an effort so that his effort parameter e is set equal to one or he doesn't and e remains equal to zero.

If both employees set e equal to zero, output equals zero with probability one. If only one employee sets $e = 1$, output is equal to $\bar{Q}$ with probability n and equal to zero with probability $1 - n$. If, instead, both employees set $e = 1$, output is equal to $\bar{Q}$ with probability m (and equal to zero with probability $(1 - m)$) where $m > n$. An employee who sets his own e equal to one bears a cost equal to d.

Supposing that the value at which the firm can sell the output $\bar{Q}$ is p and the firm could pay employees d in exchange for their effort, it would ask one of them to make an effort if

$$ np - d > 0 \qquad \text{or} \qquad n > \frac{d}{p} \tag{2.3} $$

16

If (2.3) is true, the firm also benefits from asking the second to make an effort if

$$(m - n)p > d \qquad \text{or} \qquad m - n > \frac{d}{p} \tag{2.4}$$

Together (2.3) and (2.4) imply that

$$m > 2\frac{d}{p} \tag{2.5}$$

so that the firm benefits from asking both to make an effort. Note that (2.5) could be true even if one of the two conditions (2.3) and (2.4) is false. Suppose, in particular, that (2.3) is false and (2.5) is true, which in turn requires that there be economies of scale so that $m > 2n$. Then, the firm would ask both to make an effort even though a single employee's effort would not be worthwhile.

If individual effort is not observable, the firm can provide incentives for effort only by paying a bonus w to each employee if output does equal $\bar{Q}$. If (2.3) is satisfied and (2.4) is not, the firm can achieve the same outcome as when payments in exchange for effort are feasible as long as its employees are risk neutral. The firm would contract with just one employee and offers to pay him an extra d/n if his output is indeed $\bar{Q}$. The employee would than make the effort and the firm would on average gain $np - d$.

The situation is more interesting if the firm wants both employees to make an effort. This occurs when, either (2.4) and (2.3) are both satisfied or (2.5) is satisfied but (2.3) is not. Given that the firm is now making bonus payments to two employees, the bonus w cannot exceed $p/2$. For any w, the payoffs to two selfish individuals are given by

$$
\begin{array}{lccc}
 & e = 1 & & e = 0 \\
e = 1 & mw - d, \ mw - d & & nw - d, \ nw \\
e = 0 & nw \quad , \ nw - d & & 0 \quad , \ 0
\end{array}
\tag{2.6}
$$

It is clear from (2.6) that both employees will set $e = 1$ is w is set sufficiently high. A w equal to $p/2$ may, however, be insufficient to generate individual effort. With this w, a single individual who assumes the other worker sets $e = 0$ would not set $e = 1$ unless $nw > d$ which requires that

$$n > 2\frac{d}{p} \tag{2.7}$$

Similarly, if he assumes the other employee sets $e = 1$, he would set $e = 0$ unless $(m - n)w > d$ or

$$(m - n) > 2\frac{d}{p} \tag{2.8}$$

17

Obviously (2.7) and (2.8) can both be violated even if (2.5) holds without (2.3) or if (2.3) and (2.4) hold simultaneously. In either of these cases, (2.6) is a standard Prisoner's dilemma even when w is set at its maximum allowable rate $p/2$. By dividing all entries in (2.6) by $mp/2 - d$ one obtains a game that is comparable to (1.18) where ℓ equals $\frac{2d-np}{mp-2d}$ while g equals $\frac{2d-(m-n)p}{mp-2d}$ (which is positive as long as (2.5) holds).

If

$$2d - np > 2d - (m-n)p \qquad \text{or} \qquad m > 2n \tag{2.9}$$

the resulting ℓ is bigger than g. Then, if altruism can be signaled costlessly, each individual's γ will be at least as large as what is necessary to make (1.19) hold with equality and both employees will make the effort. Since that γ equals $g/(1+\ell)$, it equals $\frac{d-(m-n)p}{(m-n)p}$. The validity of (2.4) implies this γ is positive while the failure of (2.8) implies that it is less than one. Equilibrium altruism depends on (2.9) because only if there are increasing returns to scale so that m is bigger than $2n$ do the losses to the other worker from abstaining from cooperation exceed the private benefits from defection.

Now suppose that (2.9) is satisfied but that there is a cost x of signaling one's altruism. As in Spence (1974) and Camerer (1988), this must be the cost to the altruist of an action (typically a gift) which is more expensive for a selfish individual. A selfish individual must find the cost so high that he does not gain by masquerading as an altruist and later getting the defection payoff nw. In the presence of this cost x, altruism may not emerge in equilibrium even if (2.9) is satisfied. To see this, suppose that A knows that B sets his γ to $\frac{d-(m-n)p}{(m-n)p}$. By becoming altruistic himself, A gains $mp/2 - d - x$. If this is positive, there is an equilibrium with altruism (this is the unique equilibrium when affection is chosen and demonstrated sequentially).

Suppose that, by spending some resources the firm can differentially affect the welfare of employees who like each other. In other words a firm that spends $2z$ can add ϕz to the utility of each worker if both his and his co-workers γ is at least as large as the $\bar{\gamma}$ that makes (1.19) hold with equality. If, instead, either of their γ's is lower, the firm's expenditure adds only ψz to each worker's utility. Goods that might have such a differential impact include expenditure on parties and sports events as well as changes in the job that allow the employees to have more social interactions.

The reason these expenditures can be useful is that they induce the individuals to choose high

γ's. For any w between d/m and d/n, (2.6) has a prisoner's dilemma structure because $nw - d$ is negative while nw exceeds $mw - d$ which is itself positive. Supposing that (2.9) is satisfied but $mp/2 - d - x$ is smaller than zero, the individuals remain selfish and the firm cannot get both of them to exert effort. Without spending $2z$, the firm can at most get $np - d$ by inducing the effort of just one employee.

Now consider paying both employees a bonus of $d/m + \delta$ where δ is an arbitrarily small positive number. To induce the individuals to become altruistic, z must be set so that $(mw - d - x + (\phi - \psi)z)$ which equals $(m\delta - x + (\phi - \psi)z)$ is positive. This requires that z be equal to $x/(\phi - \psi)$. Since workers can expect to get ϕz in amenities, it seems reasonable to assum ethat they they will accept a base wage that is lower by ϕz so that the total cost of friendship-enhancing activities is $2(1 - \phi)z$. Total profits are then

$$mp - 2(d + m\delta) - 2(1 - \phi)z = mp - 2(d + m\delta) - 2\frac{1 - \phi}{\phi - \psi}x$$

Given that (2.9) is satisfied, these profits are positive and bigger than $np - d$ (the profits of having only one employee exerting effort) as long as δ and $\frac{1-\phi}{\phi - \psi}$ are small. This latter ratio is small as long as ϕ is close to one (so that the z expenditures are not too wasteful) and ϕ is substantially larger than ψ (so that the expenditures are useful in promoting friendship). The expenditure of $2z$ helps the firm here because it overcomes the costs of signaling friendship.

Next, I show that the expenditure of z can be valuable even in the absence of signaling costs x as long as (2.9) is violated. The failure of (2.9) implies that an individual, say B, whose γ equals to $\bar{\gamma}$ sets $e = 1$ regardless of the sentiments of A. A is thus better off by keeping his γ equal to zero since this gives him nw instead of $mw - d$. With the expenditure of $2z$ by the firm, however, A would be willing to set his own γ equal to $\bar{\gamma}$ as long as

$$(\phi - \psi)z + mw - d \geq nw \qquad (2.10)$$

If the firm chooses z to satisfy (2.10) it thus ensures that there is a Nash equilibrium where the employees like each other. This is indeed the unique Nash equilibrium if the two individuals choose their γ's in sequence (and before they set their effort). If A chooses γ first, he can be confident that B will respond to A's altruism by setting his own γ equal to $\bar{\gamma}$. Equation (2.10) then implies that it is also in A's best interest to set γ equal to $\bar{\gamma}$.

19

I now show that the firm can benefit from setting w and z in this way. Since (2.9) is violated, (2.5) cannot be satisfied without (2.3). As a result, the firm has the option of earning $np - d$ by inducing only one individual to make the effort. Suppose that, instead, the firm sets w and z so that (2.10) is satisfied with equality. As long as the prisoner's dilemma structure of (2.6) is preserved by the choice of w, the workers then set high γ's and cooperate. Expected profits, π_z are then

$$\pi_z = m(p - 2w) - 2(1 - \phi)z = mp - 2w\left[m - \frac{1 - \phi}{\psi - \phi}(m - n)\right] - 2d\frac{1 - \phi}{\psi - \phi} \qquad (2.11)$$

Before we can evaluate the advantages of making z positive, we must determine whether the bracketed term in (2.11) is positive. If it is negative, profits are increasing in the wage paid to the employees. This means that the firm cannot do better than setting w equal to $p/2$ which leads to zero profits even with z set equal to zero. Since the firm could have made positive profits equal to $np - d$ by having only one employee, the strategy of keeping two employees with a positive z does not make sense in this case.

Now consider the case where the bracketed term is positive which requires that

$$\frac{1 - \phi}{\psi - \phi} < \frac{m}{m - n} \qquad (2.12)$$

Since profits are now declining in wages, it makes sense to set wages at the lowest level that preserves the prisoner's dilemma structure, namely slightly above d/m. I will thus assume that w equals $d/m + \delta$ and consider the limit as δ goes to zero. Using (2.11) this gives profits of

$$\pi_z = mp - 2d - \frac{2(1 - \phi)n}{(\phi - \psi)m}d - 2\delta\left[m - \frac{1 - \phi}{\psi - \phi}(m - n)\right]$$

In the limit where δ equals zero, this is greater than $np - d$ if

$$\frac{1 - \phi}{\psi - \phi} < (m - n)\frac{p}{d} - 1 \qquad (2.13)$$

So, the expenditure of z to induce the employees to become friends with each other is profitable as long as (2.12) and (2.13) are met. Both require that $\frac{1-\phi}{\phi-\psi}$ be small.

2.3. The Evidence from the Hawthorne Experiments

In the previous subsection we saw that incentives for group output in addition to the expenditure of resources to facilitate friendship may be necessary to raise productivity. In this regard,

20

the experiments at Western Electric's Hawthorne plant reported by Roethlisberger and Dickson (1939) are particularly relevant because they varied both the incentives under which groups of workers operated and their opportunity to socialize (and therefore the ease with which they could become friends). My interpretation of these experiments is that output had the largest and most sustained rise when groups were given both strong incentives and opportunities to socialize. Either the opportunity to interact or strong incentives, by themselves, had a much smaller impact.

Before being placed in the experimental conditions of the "Relay Assembly Room Experiments" the payment to the participating workers depended on the aggregate output of a group of approximately 100 operators who were located in a large room. In the experiments, 6 workers of which 5 were assemblers and one was a layout operator, were separated into a small room and paid as a function of the output of this smaller group. In the experimental conditions, a worker who produced an additional unit thus received an additional 1/5th of the group's piece rate whereas, in the initial conditions, she received only 1/100th. Not surprisingly, given that the payment per marginal unit rose, output rose as soon as the payment system was changed.[12] After this change in compensation, there were several experimental changes in the amount of rest the workers were allowed to take during the working day. During these rest pauses the workers socialized. Rest pauses were first introduced, then increased, then reduced again. The principal finding of Roethlisberger and Dickson (1939) is that output increased permanently by about 30% after these changes.[13] This change in output is attributed by them to the fact that the 5 workers became friendly with their supervisors. There were better relations between management and workers.

Roethlisberger and Dickson (1939) conducted two additional experiments. In the "Second Relay Assembly Experiment", the compensation scheme was changed and the experimental subjects were put in close proximity but the method of supervision stayed as before and there were no rest pauses. This group's output also increased but only by 12.6%.[14] Thus, the opportunity to socialize also contributed to productivity, as the model of section 3.3 predicts if one regards rest pauses as offering a disproportionately valuable opportunity for socializing to the people who like each other.

In the final experiment a group of workers originally paid on the basis of individual piecework was segregated and given the same rest pauses as in the "Relay Assembly Room Experiments".

---

[12] See what occurred in Period III on the Figure on p. 56.

[13] See Roethlisberger and Dickson (1939) p. 160.

[14] *Ibid.*, p. 129-132.

This experiment, called the "Mica Splicing Room Experiment" had the advantage that the payment system was the same in the original and the experimental conditions. Output first rose by about 15% but, unlike what was observed in the Relay Assembly Room then declined so that, by the end of the experiment it was only 4.4% higher than initially.[15] Thus, again consistent with the model of section 3.3 the incentive system also promoted productivity.

Roethlisberger and Dickson's (1939) emphasize the change in supervisory methods as explaining the change in productivity in the "Relay Assembly Room Experiments". This explanation does not seem consistent with the results of the Mica Test Room, however. Their own discussion of this experiment is more consistent with a model like the one presented in this section which emphasizes the interpersonal relations between employees. In explaining why attendance was much more regular in the later stages of the Relay Assembly Room experiment, they say it was "...more probable that the difference in attendance was due to a difference in attitude...Among the Mica operators there was no pressure for attendance except that which came from the individual's own personal situation".

In addition, "The Mica operators did not join in common social activities outside working hours. They had no "parties" similar to the gatherings of the relay assembly girls... There was no willingness to help one another...The Relay Assembly Test Room was a "group" story, the Mica Splitting Test Room was a story of "individuals"".[16] The authors own explanation is "the relation between the type of payment system to the question of morale was overlooked...Under group piecework, the operators in the Relay Assembly Test Room had a common interest under which they could organize. Under individual piece-work each girl was self sufficient; there was no need of working together".[17]

What Roethlisberger and Dickson (1939) are suggesting is that not only the rest pauses but also the group piece work encouraged the friendship among the "Relay Assembly Room" workers and that this friendship was in turn instrumental in raising output. This is important because it provides evidence for the view that friendship arises endogenously when it is in the worker's best interest rather than being simply a reaction to rest pauses and the like. To explain the increase in friendship with the model in this section, one needs to argue that the actions of the operators were

---

[15] *Ibid.*, p. 148.
[16] *Ibid.*, p. 155-6, quotes in original.
[17] *Ibid.*, p. 158.

strategically complementary. While most operators initially worked independently, there were at least two potential sources of complementarity. The first is that the layout operator had to help all assemblers. The marginal product of each assembler was thus a positive function of the layout operator's effort. Second, the assemblers found it possible to increase joint output by helping each other.

2.4. Other Evidence on Cohesiveness and Productivity

Stodgill (1972) surveys 34 studies of the connection between how much members of a group report liking each other (cohesiveness) and productivity. The results are quite mixed, about 1/3 of the studies find no relationship, 1/3 find a significantly positive relationship while 1/3 find a negative one. A more recent study that also shows a trivial correlation of output with cohesiveness is Gladstein (1984). The question I explore here is whether these mixed results are consistent with assuming that cohesiveness is equivalent to altruism. One study which does obtain both negative and positive correlations between cohesiveness and productivity and which is consistent with this view is Berkowitz (1956).

Berkowitz (1956) studied the behavior of individual experimental subjects who believed themselves to be producing output together with a co-worker (even though no co-worker actually existed). Each subject's output was supposedly used by their co-worker to make a finished product. The perceived production function thus exhibited strong complementarities between the effort of the two workers.

While first telling the subjects about their task, Berkowitz (1956) also induced favorable or unfavorable feelings for the co-workers. This was done by telling the subjects either that the co-workers were likely to be compatible with the subject or by telling them that no such compatibility existed. The subjects then briefly met their "co-workers" (who were just confederates of the experimenter). Those told that their co-workers were compatible did later express much greater liking for their co-worker than those told that they were incompatible.

The subjects then carried out their assigned task in isolation. During this period they received messages which they were told had been written by their co-workers. The key result of the study is that the subjects who expressed that they liked their co-workers were much more likely to raise their effort when these messages said that the co-worker needed more inputs. They were also more

23

likely to reduce their output if the messages said that the co-worker was either tired or inundated with materials. Thus, workers who express that they like their co-workers in questionnaires also choose their output taking into account the desires of their co-workers.

What is less clear is whether the theory correctly predicts the sentiments of the subjects. The problem is that we do not know what the subjects believed to be their payoff. Insofar as they felt that there was some benefit in producing a high level of final output, each subject's effort was a strategic complement with the effort of their co-workers. Thus, the theory predicts that the subjects will tend to like their co-workers. The theory can also explain the lack of affection between the "incompatible" co-workers if the subjects felt that their personal payoffs had a prisoners's dilemma structure as in Section 1.2. In that setup, there is no benefit from liking someone that one does not expect to be liked by.

3. Relations of Authority

In this section, I study the relationship between employees in authority with their subordinates. I will argue that, unlike what is predicted by the models of collusion in hierarchical agency (Tirole (1986), Holmström and Milgrom (1990)), the current model predicts relatively little ganging-up between superiors and their subordinates against the firm as a whole. The reason is that the strategic variables that they control do not tend to be strategic substitutes.

I consider three models. The first is essentially the model of Tirole (1986) which deals with a single employee whose supervisor's only role is to report on the employees' effort. The second involves two different employees. The third involves a relationship of some authority but, unlike the previous two models, it involves no ability by the "supervisor" to affect the employee's pay. This last example is based on a real-life interaction described by Crozier. The "supervisors" Crozier interviewed expressed that they liked their fellow employees even though they were not well liked in return. This seems hard to reconcile with the view that friendship is an automatic response to common experiences but, I will argue, it fits well with my model of rational altruism.

3.1. The Supervisor as Performance Evaluator of a Single Employee

In this subsection, I consider the problem posed by Tirole (1986) of a supervisor whose sole role is to report on a single subordinate's productivity. The subordinate's output is given by

$$Q_A = \epsilon_A + a \tag{3.1}$$

where ϵ_A is a productivity parameter outside the agent's control while a represents his effort. The cost of effort is $e(a)$. The parameter ϵ_A can take two values ϵ^h and ϵ^l with $\epsilon^h > \epsilon^l$. The employee observes the realization of ϵ_A before choosing his effort. The supervisor sometimes observes the realization of ϵ_A and sometimes doesn't. When he does, he can convey this information to top management in a manner that is credible. In other words, when the firm is given this information it can verify its validity. The supervisor thus has only one opportunity for lying, he can say that he observed nothing when he actually knows the state of technology.

Both the supervisor and the employee are risk averse. The latter's utility is given by $EU(w - g(e))$ where E is the expectations operator, U is a concave function and p is the employee's wage. Thus, as Tirole (1986) shows, the firm's optimal contract without collusion has the supervisor

receiving a constant wage and always making a truthful report. Rather than deriving the rest of the optimal contract, I simply describe intuitively its salient features. For details, the reader is referred to Tirole's (1986) Appendix. Following Tirole (1986), let 1 and 4 describe the two states where the supervisor does observe ϵ_A. Then, if the supervisor is truthful, effort and wage of the subordinate should be the same in these states. Denote these by w_1 and e_1.

Now consider states 2 and 3 where the supervisor does not observe the outcome and let 3 be the state where ϵ_A is high. If the wage were also w_1 in these two states, the subordinate would always claim that the state is low so that he can make a low effort in the high state. To induce him to produce more when the state is high, he must be paid more in state 3 than in state 2 so that $w_3 > w_2$. And, to make the marginal utility of the individual vary as little as possible, the contract sets $w_3 > w_1 > w_2$. Finally, effort in state 3 is the same as in states 1 and 4 because the firm has nothing to gain from distorting effort in state 3.[18]

Because $w_3 > w_1$ the subordinate would gain if the superior claimed not to know the state when $\epsilon_A = \epsilon^h$. This would allow the subordinate to receive w_3 instead of w_1. The subordinate is thus willing to pay something to the superior to distort his report. I now show that, in spite of this benefit from collusion, endogenous altruism is not sufficient to induce cooperation between superior and subordinate.

First, the subordinate has nothing to gain from feeling altruism towards his superior. Even if the superior is paid a flat wage, the subordinate's affection does not lead him to change his report and is thus of no benefit to the subordinate. By being altruistic towards the subordinate, the superior makes the subordinate better off (since the superior will dissemble in the state where he observes ϵ^h). However, this provides no benefits to the superior. Thus, the superior will not choose to become altruistic towards the subordinate.

This interaction is actually a special case of the following

$$
\begin{array}{ccc}
 & C & D \\
C & f,\ z+g & 0,\ z+g \\
D & f,\ z & 0,\ z
\end{array}
\tag{3.2}
$$

Each player benefits from the other playing C but has no control over his own outcome. The result is that one's strategy depends on one's affection for the other player but not on his affection for

---

[18] It does stand to gain from lowering effort in state 2 below the first best because this underproduction in the low state, accompanied by a corresponding reduction in w_2 reduces the attractiveness of claiming that $\epsilon_A = \epsilon^l$ when $\epsilon = \epsilon^h$.

oneself or even his action. There is then no private benefit from altruism. What I have shown is that, if cooperation in the workplace arises only as a result of self interested altruism, it will not arise in settings such as Tirole (1986), even though cooperation is beneficial in such settings.

Tirole (1986) provides several examples of superiors distorting the information they pass along to the central office and regards these as suggesting that collusion of the form he models is quite common. For instance, he cites Dalton (1959 pp. 80-85) who illustrates several cases where industrial accidents are not reported. In the most vivid case the foreman took a worker who had broken several toes to a private physician and paid for his medical bills out of his own pocket. Apparently, such accidents could have been avoided if the workers had worn safety shoes but the workers regarded these as cumbersome. Thus, it is possible to read this case as one where the workers somehow bribe the foreman to allow them to wear regular shoes. But, it is just as possible that the foreman himself expects higher productivity (and thus more personal income) from workers who wear regular shoes. It might then be in his own self-interest (leaving aside any implicit contract with the workers) to evade the rule on safety shoes even if this means he has to pay certain medical bills when accidents do occur.

Crozier (1964) provides additional evidence on what motivates supervisors to lie - and lie they do. When talking of the immediate supervisors (or section chiefs) he says "... they are likely to bias the information they give in order to get the maximum of material resources and personal favors with which to run their sections smoothly."[19] This lying is interpreted as self interested by the workers. In response to the question "whether the supervisor would go to bat for employees", a worker responds: "Supervisors defend us inasmuch as they think it will help them. If not they would eventually trample on the employees. Our own section chief is very self seeking".[20] The perception that supervisors do not feel warmly toward their subordinates is consistent with the supervisors' own statements. Crozier reports that supervisors "do not have high regard for production workers, whom they consider neglectful, irresponsible and careless".[21]

The picture that emerges is consistent with my model in that the lying is certainly not motivated by affection. What is less clear is whether, and in what sense, these self-seeking supervisors are colluding with the workers. These examples certainly do not involve direct *quid-pro-quos* where

---

[19] *Ibid.*, p. 45.
[20] *Ibid.*, p. 42.
[21] *Ibid.*, p. 92.

workers induce their supervisors to lie in exchange for personal favors.[22] On the other hand, supervisors do appear to feel that their sections would run more smoothly with more resources and this suggests that they are in a position to increase output when given more resources. Whether these resources are used to "pacify" the workers and whether, if they do, this should be viewed as "collusive" remains an open question.

In the formal model of this subsection supervisors neither gain nor lose from liking their employees. There are, however, two reasons why supervisors will generally be better off without altruism for their subordinates. The first can be seen by modifying the model slightly so that the supervisor is given some small bonus b when he actually reveals the state of nature. Since the supervisor's information is verifiable, such a contract is feasible. Now, altruism towards his subordinate makes the supervisor worse off because, in state 1 he would claim not to know the state and lose his bonus.

The second reason for the supervisor to lose from his altruism arises when instead of having one subordinate, the supervisor has several. In this case, one important role for the supervisor is to maintain incentives for his subordinates by inducing them to compete with each other as in section 4. I show in the next subsection that this becomes more difficult when the supervisor starts liking one subordinate more than he likes the others.

3.2. The Supervisor as Performance Evaluator of Two Employees

Suppose that there are two subordinates A and B and that, as in section 4.1, the supervisor sometimes knows the true state of nature while he sometimes doesn't. In particular, let the supervisor sometimes know the ranking of the two subordinate's outputs while, in other cases, he and the firm as a whole, know only their sum. The output of A is again given by (4.1) while that of B, Q_B, is, analogously given by $\epsilon_B + b$ where ϵ_A and ϵ_B are positively correlated random variables.

With probability π, the firm and the supervisor observe only $Q_a + Q_b$ so that they must compensate them as in section 2. With probability $1 - \pi$, the supervisor knows, and can prove to top management, which worker produced the most. To provide incentives for effort, it then makes sense to penalize the worker who has produced less and give a bonus to the worker who

---

[22] Dalton (1959) p. 58 does provide examples of cliques which lie for each other in order to ensure the promotion of all the members of the clique. Unlike what is implied by the Tirole (1986) model, the verbal description of this interaction does seem consistent with strategic complementarity.

has produced more. This is generally better, for insurance purposes, than simply paying more to the more productive worker.[23] I now ignore employee friendship to focus on the altruism of the supervisor for one employee. Incorporating friendship between the subordinates would not affect the analysis as long as that friendship was not so intense as to destroy the incentives from the tournament component of compensation.

If the supervisor starts liking one person, he ceases to reveal his knowledge in the cases where his favorite is second. This eliminates the incentive effect of the contract on the employee who is not the favorite (since he never officially wins). The favorite still has some incentive to make an effort from the tournament component since he only collects the prize when his victory can be verified. Nonetheless, the reduction in the effort of the non-favorite employee also reduces the favorite's effort (as I show below in Section 4). So, output falls. If the supervisor's income depends positively on output, he is thus worse off by liking one employee. Not surprisingly, favoritism can be quite detrimental for supervisors.

3.3. The Team Leader

In this subsection, I will consider an authority relation based on Crozier's (1964) account of work in a French clerical agency. In this agency, the actual work was done by teams of four people whose job was to process customer's requests. The first member of the team checked the customers records. The next two processed the requests that were approved by the first. The final member of the team checked the whole transaction. The first member was the team leader in the sense that she set the pace even though she did not have any formal authority nor ability to set compensation.[24]

What makes this relationship interesting from our point of view is the difference in affective attitudes among the members of these groups. The employees had very few ties of affection with their workmates. As Cozier says: "Only 20% feel positively about their workmates as potential friends"[25] And, most of these feelings of affection flowed from the team leaders to the other members

---

[23] This can be seen as follows. Suppose that we are considering a state of nature where, in the absence of information, they both get w. Suppose that the contract then says that, with information on rank, the most productive worker gets $w + z$ while the other keeps getting w. With concave utility, one can always improve on this by offering a slightly higher wage in the absence of information and a slightly lower wage in the case where the employee comes in second. The individual, and the firm, are indifferent to small perturbations of this sort (because utility is locally linear). This change allows the difference in utility between the winner and loser (which is what generates the effort) to remain the same at a substantially lower wage when the employee wins. This lowers utility in those states of nature. Nevertheless, expected utility rises if average income is kept constant by increasing the income the individual gets in the states with no information. The reason is that, with concave utility the marginal utility of income is higher in those states.

[24] See Crozier (1964) p. 18.

[25] Ibid., 35.

of the team: "The middle-class girls who made positive comments about their workmates as possible friends were team leaders (half of them) and a few senior employees from the special workroom".[26]

I now present a model of a relationship between a team leader and a single other team member that is loosely based on Crozier's account. I show that, in this model, the team leader will choose to become altruistic towards the other team member but not otherwise. Thus, one nice feature of the model is that it can explain situations where only one person chooses to like the other.

In this model, the team gets paid as a function of their joint output Q; each member is paid vQ. The payment v per unit of output should not be regarded as a piece rate which is set before production takes place. Rather, it is an *ex post* payment which depends on whether the employees delivered large quantities of output when it was important. Thus, teams who perform well in this regard might get promoted, might be favored when it comes to allocating better office equipment, etc.

The leader sets the pace in the sense that she picks Q. The other team member (whom I will term the follower) has a more restricted choice. He can agree to the pace Q or complain. In practice complaints about the behavior of co-workers (including the leader) were common and they led to an investigation into the appropriate behavior of the co-worker.

The leader's cost per document depends on idyosincratic factors as well as the features of the document that are being handled. I assume that the leader's total cost is $\frac{Q^2}{2} - zQ$ where z is a variable known only to the leader. Thus, a leader would like to set Q so that it maximizes

$$(v + z)Q - Q^2$$

and would thus choose

$$Q = \frac{v + z}{2} \tag{3.3}$$

assuming this is positive. I assume that the mean of $(v + z)$, $(\bar{v} + \bar{z})$, is positive so that the leader's optimal output is positive for the mean realization.

The follower's cost is given instead by $Q^2 - rQ$ where r is common knowledge. I will make the key assumption that the follower does not know v and z. The idea is that the leader is in a better position to know when production is important. While this asymmetric information

---

[26] Ibid., 36.

is not made explicit in Crozier (1964) he does say that the role of the leader falls to the most experienced operator and that, unlike what the official guidelines require, this job is not rotated among employees. To ensure that the follower wants to produce positive amounts in the absence of any information about v except for its mean, we must have $\bar{v} + r > 0$.

For simplicity, I assume that a complaint by the follower gives him a (suitably normalized) payoff of zero. It might be more realistic to assume that the follower's payoff depends on the results of the ensuing investigation. I do not pursue this because the follower's response is the same in both cases: he should complain whenever the output Q chosen by the leader is above some threshold level. The reason this is optimal for the follower is that high values of Q suggest that z is high and thus high effort on the follower's part is only in the leader's interest. To see this assume that z and v are independently distributed with normal distributions defined by $N(\bar{z}, \sigma_z^2)$ and $N(\bar{v}, \sigma_v^2)$ respectively. Then, if the leader chooses Q using (3.3), the follower's expectation of v given Q is $\bar{v} + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_z^2}(Q - \bar{v} - \bar{z})$. Therefore, the followers expected payoff if he goes along with this choice of Q is

$$\left[\bar{v} + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_z^2} \left(Q - \frac{\bar{v} + \bar{z}}{2} \right) \right] Q - Q^2 + rQ \qquad (3.4)$$

Assuming the leader picks Q as in (3.3) and that the follower gets zero when he complains, the follower's cutoff is that value of Q (other than zero) that makes the expression in (3.4) equal to zero. Thus the cutoff $Q^{\max}$ satisfies

$$Q^{\max} = \bar{v} + r + \frac{\sigma_v^2}{\sigma_z^2} \left(\frac{\bar{v} - \bar{z}}{2} + r \right) \qquad (3.5)$$

Given any cutoff $Q^{\max}$ the leader's best choice is to pick Q as in (3.3) whenever this is below $Q^{\max}$ and pick $Q^{\max}$ otherwise. Thus, in equilibrium, the cutoff is given by (3.5) with output given by the minimum of the quantity in (3.3) and that in (3.5).[27]

Consider now what happens when the follower chooses to become altruistic towards the leader. This leads the follower to pick a higher level for the cutoff $Q^{\max}$. This is clearly in the leader's best interest. There is now a larger set of realizations for v and z such that he chooses Q as in (3.3). However, the follower does not benefit from this change in emotion. Indeed, while small changes in the cutoff have no effect on follower welfare, large increases are strictly deleterious. The reason

---

[27] This ignores negative outputs. Obviously, if (3.3) gives a negative Q, the optimal Q is zero.

is that output doesn't change except in states of nature where the follower would have expected to be better off with an output of Q^{max}. The follower thus has no incentive to feel altruistic towards the leader.

Finally, consider an increase in the affection of the leader for the follower. The leader now chooses Q to maximize

$$(1 + \gamma_A)(vQ - Q^2) + (z + \gamma_A r)Q$$

so that, if he were unconstrained, he would pick output to ensure that

$$Q = \frac{(1 + \gamma_A)v + z + \gamma_A r}{2(1 + \gamma_A)} \tag{3.6}$$

If the follower believes that Q is chosen in this way (so that he believes in the affection of the leader), his expected value of v given Q is

$$\bar{v} + \frac{\sigma_v^2}{\sigma_v^2 + (\sigma_z^2/(1 + \gamma_A)^2} \left(Q - \frac{\bar{v}}{2} - \frac{\bar{z} + \gamma_A r}{2(1 + \gamma_A)} \right)$$

Thus, his payoff if he goes along with the choice of Q given by (3.6) is

$$\left[\bar{v} + \frac{\sigma_v^2}{\sigma_v^2 + (\sigma_z^2/(1 + \gamma_A)^2} \left(Q - \frac{\bar{v}}{2} - \frac{\bar{z} + \gamma_A r}{2(1 + \gamma_A)} \right) \right] Q - Q^2 + rQ$$

Equating this to zero and solving gives the new value of the cutoff which, after some manipulation to make it comparable to (3.5) is given by

$$Q^{\text{max}} = (\bar{v} + r) \left(1 + \frac{\gamma_A \sigma_v^2}{2\sigma_z^2} \right) + \frac{\sigma_v^2}{\sigma_z^2} \left(\frac{\bar{v} - z}{2} + r \right) \tag{3.7}$$

This expression is increasing in γ_A as long as $(\bar{v} + r)$ is positive, which is required for the follower to want to produce positive output in the absence of any information.

This implies that the leader benefits from small quantities of affection for the follower. Because (3.3) maximizes leader welfare, a small increase in γ_A has no effect on her welfare when she is not constrained by the cutoff. On the other hand, an increase in γ_A does raise the cutoff and this makes her strictly better off. Intuitively, the leader benefits from her affection because this affection leads the follower to trust her more and, as a result, they both increase their effort when the payoff from doing so (v) is high.

4. Relative Performance Evaluation

Suppose A and B independently carry out similar tasks but that the outcomes of the two tasks are correlated. Thus, if individual i's output is denoted by Q_i, the two outputs are given by

$$Q_A = h(a, \epsilon_A) \qquad Q_B = h(b, \epsilon_B) \tag{4.1}$$

where the h function is increasing in both arguments while ϵ_A and ϵ_B are positively correlated random variables outside of the control of the individuals. Then, assuming the firm does not observe the random disturbances and the two employees are selfish, it makes sense to offer them a compensation scheme which makes their own wage depend not only on their own performance, but also on the other agent's performance (Lazear and Rosen (1981), Holmström (1982), Green and Stokey (1983)). The reason is that the other agent's performance conveys information about the common component of the two ϵ's. Thus, the compensation of A is given by $C(Q_A, Q_b)$ where the C function is increasing in the first argument and decreasing in the second. Similarly, B's compensation is given by $C(Q_B, Q_A)$.

Let $e(a)$ and $e(b)$ be, once again, the private costs of effort. The resulting payoff for A is then

$$\alpha = C\Big(h(a, \epsilon_A), h(b, \epsilon_B)\Big) - e(a) \tag{4.2}$$

and similarly for β.

It is apparent from this formulation that α_b (and β_a) are both negative so that employees benefit each other when they cut down on their effort. As shown by Holmström and Milgrom (1990), collusion between employees thus reduces effort. But, will altruism arise naturally? We can be sure that it will arise if α_{ab} is positive - which requires that C_{12} be positive - but this need not be part of the optimal contract.

To illustrate the ambiguities involved, I analyze a particularly simple contract that has been shown by Lazear and Rosen (1981) and Green and Stokey (1983) to have desirable properties in this setting. With this contract each employee's payment depends only on the rank of his output.[28] Let the worker with the highest output be paid W_2, while the one with the lower output is paid W_1 where $W_1 < W_2$. Specializing further to the case where the h function is given by the sum of

---

[28] Note that such a tournament does indeed make payments insensitive to the other employees output when the difference between the two outputs is large.

33

the effort and the disturbance (so that $h(a, \epsilon_A) = a + \epsilon_A$) the payoffs to A and B are

$$\alpha = W_1 + (W_2 - W_1)G(a - b) - e(a)$$
$$\beta = W_1 - (W_2 - W_1)G(a - b) - e(b)$$

(4.3)

where G is the c.d.f. of $\epsilon_A - \epsilon_B$.

At the Nash equilibrium without altruism we have

$$\alpha_a = (W_2 - W_1)g - e' = 0$$

(4.4)

Thus α_{ab} has the same sign as $-g'$. If the distribution G is symmetric, differentiable and has a mode at zero, the symmetric Nash equilibrium occurs at a point where $g' = 0$.[29] Thus, there is no incentive for small changes in γ at this symmetric equilibrium. The intuition for this result is that, in this case, the slope of expected compensation with respect to effort is highest at the Nash equilibrium so that it does not change for small changes in the other person's action. This still leaves open the question whether larger changes in γ are attractive. If G is twice differentiable, the second derivative of g with respect to its argument at its unique mode is negative. This implies that relatively large drops in b lower g and, if A behaves as in (4.4), lower a as well. Thus, B gains something from relatively large increases in γ_B (since these lower b and lead to reductions in a). Of course, these large reductions in b also lead to direct costs for B so that one cannot be sure that B benefits from his altruism. Intuitively, what goes on is that large reductions in b not only raise the probability that A will win but also lower the sensitivity of the probability of winning to A's effort. This is what reduces a.

To gain further insight into what determines whether B benefits, I consider a Taylor expansion of G and e around the values that their arguments take in the "selfish" Nash equilibrium. These are 0 and $a^N = b^N$ respectively. An increase in γ_B lowers b (since α_b is negative). Using the Taylor approximation, the resulting value of e is then

$$e(b) = e(b^N) + e'(b - b^N) + \frac{e''}{2}(b - b^N)^2$$

(4.5)

and similarly for $e(a)$. In the new equilibrium, b will be below a so that $a - b$ is positive. The Taylor approximation of G can thus be written as

$$G(a - b) = G(0) + g(a - b) + \frac{g'}{2}(a - b)^2 + \frac{g''}{6}(a - b)^3$$

(4.6)

---

[29] These conditions on the distributions are obviously met when the distribution of each ϵ is normal as in Lazear and Rosen (1981).

where g' is equal to zero. Using these approximations, the first order condition (4.4) for A, which remains selfish, becomes

$$(W_2 - W_1)\left(g + \frac{g''}{2}(a - b)^2\right) - e' - e''(a - a^N) = 0$$

Using (4.4), this implies

$$a - a^N = \frac{g''}{2e''}(W_2 - W_1)(a - b)^2 \tag{4.7}$$

Because g'' is negative while e'' is positive, this equation says that, if b is changed sufficiently from its selfish Nash equilibrium value, a declines. Using (4.3), (4.5) and (4.6), the resulting difference between β and its value at the selfish Nash equilibrium is

$$-(W_2 - W_1)\left(g(a - b) + \frac{g''}{6}(a - b)^3\right) - e'(b - b^N) - \frac{e''}{2}(b - b^N)^2$$

Given that these expressions are evaluated at the Nash equilibrium (4.4), and that $(b - b^N)$ is equal to $(a - a^N - a + b)$, this difference equals

$$e'(a^N - a) - (W_2 - W_1)\frac{g''}{6}(a - b)^3 - \frac{e''}{2}(a - a^N - [a - b])^2 \tag{4.8}$$

As B raises γ_B, b falls so that $(a - b)$ becomes positive and, given (4.7) a falls. Thus the first two terms in (4.8) are positive (meaning that B gains) while the last term is negative. However, as long as e', or equivalently $(W_2 - W_1)g$, is large relative to g'' and e'', the positive terms dominate and B gains from his altruism. Similarly, as we shrink e'' keeping e' and g''/e'' constant, we can be sure that B benefits from liking A.

Tournaments also promote altruism and cooperation in the case where the ϵ's have independent uniform distributions. While this case may seem rather special, it illustrates another reason why the sensitivity of A's probability of winning to his effort falls when b falls. The reason is that reductions in b create a range of realizations for ϵ_B such that A wins whether he changes a slightly or not. When ϵ_A and ϵ_B are independently distributed, with density function f, G becomes

$$G(a - b) = \int_{-\infty}^{\infty}\int_{b-a+\epsilon_B}^{+\infty} f(\epsilon_A)f(\epsilon_B)d\epsilon_A d\epsilon_B$$

so that g is

$$g(a - b) = \int_{-\infty}^{+\infty} f(b - a + \epsilon_B)f(\epsilon_B)d\epsilon_B$$

35

In the case where f is the uniform distribution between 0 and 1, and b is smaller than a, this equals $(1 - (a - b))$ so that, from (4.4), α_{ab} is positive.[30] Thus, it is in B's interest to raise γ_B above zero. This then leads to a decrease in effort. The model thus explains why, in some circumstances, altruism and friendship between employees rises endogenously and hurts firm profitability.

The model also helps explain why firms do not resort as much to tournaments as one would expect from Lazear and Rosen (1981). Indeed, Lazear (1989) provides several examples where firms appear to shy away from having people who are near each other compete even though such competition would appear optimal.[31] From the point of view of firms, such competition is not attractive in my setting because individuals respond by liking each other and thereby blunting the competition. It might be argued that some competition is still likely to remain in equilibrium. One problem, however, is that the personalities involved affect the degree of friendship and altruism. Personality matters if it affects α, β, or the direct benefits from the friendship itself. Thus, firms who put in place tournaments subject individuals to additional risk since they make individual's outcomes very sensitive to the personality of the worker with whom they are matched. This risk reduces the *ex ante* welfare of workers, and tournament-using firms who wish to attract workers must therefore pay higher wages.

---

[30] Note that g is actually equal to $(1 - |a - b|)$ so the sign of α_{ab} depends on whether a is bigger than or smaller than b. However, as required for B to benefit, if B starts liking A, b falls so $a - b$ becomes positive and a responds by lowering a.

[31] Lazear's (1989) theory of this stresses the opposite of altruism. He notes that employees who compete with each other are likely to sabotage each other's work, to the detriment of their employer.

5. Sequential Moves

So far I have generally assumed that A and B select their actions a and b simultaneously. While somewhat artificial, this modelling device does ensure that A and B are treated symmetrically. In this section, I briefly consider the case where A always makes his move first. I do this both to study the robustness of my conclusions and to relate this work to that based on Becker (1974a).

Consider the continuous action model of section 2 and assume that A chooses a first. B then makes his choice of b with full knowledge of a. For any value of γ_B (including zero), the first order condition which gives the resulting value of b is (1.8). Since a is known at this time, (1.8) gives b as a function of a and γ_B. Denote this function by $\Phi(a, \gamma_B)$. We know the local behavior of this function from the second line of equation (1.11), which can be rewritten as

$$db = -\frac{\beta_{ab} + \gamma_b \alpha_{ab}}{\beta_{bb} + \gamma_b \alpha_{bb}} da - \frac{\alpha_b}{\beta_{bb} + \gamma_b \alpha_{bb}} d\gamma_b \qquad (5.1)$$

From (5.1) it follows that, if a and b are strategic complements for both players, b rises when a does. Also, given the normalization that α_b is positive, b rises with γ_B.

A's problem is different; he chooses a knowing that B will choose b using the function Φ. Thus, a selfish A would set a such that

$$\alpha_a + \alpha_b \Phi_a = 0 \qquad (5.2)$$

It should now be apparent that, because b does not depend on γ_A but only on a itself, A has nothing to gain by being altruistic. In other words, he achieves the highest value of α by keeping γ_A equal to zero. The reason is that altruism has value by influencing the other players' perception of what one is likely to do. But, in the case where B moves last, B's perception doesn't matter, since he already knows the action a.

This result, that any altruism that occurs in equilibrium must be B's, accords well with the timing of moves in Becker (1974a, 1976) where the altruist moves last. It therefore suggests that Becker's order of moves need not be an assumption but a conclusion from letting altruism be endogenous. This result does hinge, however, on a predetermined order of moves. Because the realism of this is questionable, and because it has such a pronounced effect on the pattern of affection, I concentrated mostly on simultaneous moves.

I now study whether B benefits from becoming altruistic i.e., whether $d\beta/d\gamma_B$ is positive.

Using (5.1) and (1.8), $d\beta$ is given by

$$d\beta = \beta_a da + \beta_b db = \left(\beta_a + \gamma_B \alpha_b \frac{\beta_{ab} + \gamma_b \alpha_{ab}}{\beta_{bb} + \gamma_b \alpha_{bb}} \right) da + \frac{\gamma_B \alpha_b^2}{\beta_{bb} + \gamma_b \alpha_{bb}} d\gamma_B \qquad (5.3)$$

The second term represents the "direct" effect which is negative given the second order conditions. I will consider the case where γ_B is small so that this term can be neglected. The first term then has the same sign as da. To obtain da, one must differentiate (5.2) with respect to a, b and γ_B. This raises the question of how Φ_a, which is the coefficient of da in (5.2), responds to changes in these variables. It is apparent from (5.2) that Φ_a depends not only on γ_B but also on b itself. Moreover, the extent of this dependence is a function of the third partial derivatives of α and β. Thus, because the sign of these partial derivatives affects the desirability of altruism, strategic complementarity is clearly not enough. I thus proceed to show only that strategic complementarity plays *some* role in generating altruism. To do this, I assume that α and β are quadratic so that db/da varies only with γ_B. The derivative db/da with respect to γ_B, $\Phi_{a\gamma_B}$ is then

$$\Phi_{a\gamma_B} = -\frac{\alpha_{ab} + \alpha_{bb}\Phi_a}{\beta_{bb} + \gamma_b \alpha_{bb}} \qquad (5.4)$$

The second order conditions imply that the denominator of (5.4) is negative. Thus, $\Phi_{a\gamma_B}$ is positive as long as α_{ab} exceeds $-\alpha_{bb}\Phi_a$. In the normal case where α_{bb} is negative, this requires a strong strategic complementarity.

Ignoring the dependence of Φ_a on b, differentiation of (5.2) gives the following expression for da

$$da = -\frac{(\alpha_{ab} + \Phi_a \alpha_{bb})db + \Phi_{a\gamma_B} d\gamma_B}{\alpha_{aa} + \Phi_a \alpha_{ab} + \alpha_b \Phi_{aa}} \qquad (5.5)$$

Combining (5.1) and (5.5), the change in a when γ_B is small is

$$da = (-\alpha_b \Phi_{a\gamma_B} \beta_{bb} + \alpha_b (\alpha_{ab} + \Phi_a \alpha_{bb}) / D' \qquad (5.6)$$

where

$$D' = (\alpha_{aa} + \alpha_{ab}\Phi_a)\beta_{bb} - (\alpha_{ab} + \alpha_{bb}\Phi_a)\beta_a b$$

The denominator D' has a generally ambiguous sign. As before, this sign can be made determinate if one assumes the system is stable. In particular, suppose that B always chooses b to

maximize his utility (so that b satisfies (1.8)) but that A has to grope towards his optimal a.[32] Assuming that the change in a has the same sign as α_a, the system is stable when D' is positive.[33]

It is apparent from (5.4) that, with a positive D', the expression in (5.6) is positive whenever the strategic complementarities are sufficiently strong to make $\Phi_{a\gamma_B}$ positive. Then, both terms in the numerator of (5.6) are positive so that a increases with γ_B. This in turn implies that increasing γ_B from zero is attractive since the increase in β then has the same sign as the change in a.

I have now established that strategic complementarities are also important in generating altruism in sequential move games. Before closing this section, I briefly discuss the application of these ideas to the "merit good" game of Becker (1991) and to the "transfer game" of Becker (1974a). In the former, the first mover (the child) picks the level of a "merit good" such as studying hard or visiting often. After this, the parent, whose utility is increasing in the amount of the merit good picked by the child, chooses how much money to transfer to the child. Becker (1991) shows that children of altruistic parents will tend to pick relatively high values of the merit good if, from the point of view of the parent, transfers and the merit good are complements, *i.e.*, if the marginal utility of transfers goes up when the child behaves well. What my analysis shows is that precisely in this case where altruism leads the kid to behave well, altruism is in the parent's self interest.

In the "transfer game" of Becker (1974a) the action taken by the first mover A affects only income and, in particular, it affects the second mover's wealth. The second mover, B, then decides how much income to transfer to A. To determine whether altruism is good for B in this game, we must determine whether the resulting increase in transfers to A leads A to generate more income for B.

I now argue that the existence of this strategic complementarity depends on whether A's action generates mostly income controlled by A or mostly income controlled by B. Whoever controls the income later decides how it is to be spent. Suppose first that A's action, a, requires costly effort to generate income only for himself and that b represents the fraction of B's income that is transferred to a. An increase in b then lowers a since it lowers A's marginal utility of income. The actions are

---

[32] This makes some sense since A can't be sure what b will be, while B does know a.

[33] To see this, note that (5.2) then implies that da/dt is proportional to

$$(\alpha_{aa} + \alpha_{ab}\Phi_a)(a - a^*) + (\alpha_{ab} + \alpha_{bb}\Phi_a)(b - b_*)$$

and use (5.1) to determine $(b - b^*)$ from $(a - a^*)$.

strategic substitutes and B loses from his altruism.[34]

Now suppose, as is implicit in Becker (1976), that A's action generates only income under the control of B. Then, letting b be positive rather than zero raises a because now A gets at least a fraction of the income generated by his effort. Thus, in this case, a and b are strategic complements for A, and B does benefit from his altruism. His altruism leads A to produce and he gets to keep a fraction of this income.

6. Conclusions

This paper has shown that altruism can be rational and that this can explain certain observations from industrial settings. The central idea is that strategic complementarity breeds affection. By becoming altruistic towards you, I am led to change my behavior. If the resulting changes in actions induce you to change your own actions in a way that benefits me, my becoming altruistic is smart. This idea, while extremely simple, is rich in refutable implications. This richness comes from the fact that the theory predicts both the emotions that people have and their actions. Insofar as emotions can be measured, there are thus more implications than in theories that predict actions alone. Here, I have only focused on measurements of emotions that are based on self-reports and have found some concordance between these and the emotions predicted by the model. To be more convincing, this evidence would have to be supplemented with evidence from physical and chemical changes which are supposed to accompany changes in emotions.

In this concluding section, I deal with two issues. The first is a comparison of my theory of altruism with Homans' (1950) idea that people tend to like (or feel altruistic towards) those from which they receive large rewards at low costs. The second is the applicability of the theory to other settings. I will argue that many of the findings that have led Homans (1950) and others to embrace this idea are also consistent with my own. In addition, this traditional theory is less consistent than my own with the observed features of authority relations.

Using anthropological data from the Polynesian island of Tikopia and data from the Hawthorne experiments as evidence, Homans (1950) stresses that the friendliness among co-workers is much

---

[34] One interesting feature of this example is that the outcome is Pareto Optimal (so that the "rotten kid theorem" holds) whether B is altruistic or not. The example is not wholly in the spirit of Becker (1974a) because a affects only A's income. However, the conclusion that a and b are strategic substitutes for A so that altruism is bad for B extends also to the case where the effort a also raises the income controlled by B as long as this effect is sufficiently small. Pareto efficiency of A's effort is still assured by the "rotten kid theorem" as long as A's utility is linear in income and B's altruism is sufficiently intense.

more pervasive than between workers and their supervisors. In the island of Tikopia there is much more friendliness between brothers (who work together in the field) than between sons and fathers (where the latter are the central authority figure). This seems problematic for the view that friendliness depends only on the frequency of interaction or even on the happiness of the outcomes. Indeed, Homans' (1950) is led to posit that friendliness from repeated interaction arises only when the participants are relatively equal, or, in his words "when no one originates the interaction with much greater frequency than the others" (p. 243). One advantage of my theory is that this difference in affective attitudes is explained endogenously as one can see by comparing Sections 2 and 3.

Moreover, my theory can explain *why* people like individuals who give them large rewards (such as people who are competent or attractive). Explaining why people feel altruistic towards those that give them large rewards might actually be regarded as something of a challenge since it may not be apparent what effect altruism has on one's own rewards from the relationship. This mystery dissipates once one realizes that my theory implies that liking someone raises the likelihood that one will be liked in return (even though there is no formal *quid pro quo*).

That affection tends, endogenously, to flow towards those that like one already can be seen in the prisoner's dilemma case of Section 1.2. There, positive outcomes for *either* individual are sometimes obtained only when *both* individuals like the other. In this case, an individual has more to gain from liking someone that already feels affection for him than from proffering his affection to an unknown third party. This has two consequences. First it explains why feelings of altruism are often symmetric, as is required by the fairness theory of Rabin (1991). Second, it implies that an individual who feels altruistic towards somebody whose company benefits him may actually be rational; he is raising the odds of continuing to profit from the relationship.

What is more, the notion that one raises the probability of being liked by liking someone has plenty of empirical support (See Aronson (1984)). So, the idea that people benefit from being altruistic towards those they benefit from interacting with is, itself, already proven by the existing experimental evidence.

In this paper, I have studied interactions only within the workplace. While affection certainly also takes place elsewhere, I have focused on the workplace for several reasons. First, firms create a

large degree of variation in the setting within which people interact. This allows one to study how relationships differ in different settings. Second, the benefits people extract from the workplace are relatively tangible. By contrast, the benefits one extracts from one's social friends and family members are not only more subjective, they are also more likely to be unconscious. Nonetheless, extensions of the model to relationships among family members and friends seems desirable. The endogenous creation of altruism within the family would seem to be a useful complement both to Becker's (1974a and 1976) discussion of the consequences of altruism and to Becker's (1991) insight that people who like each other are more likely to marry.

Another potential application of the model is to the ties of friendship that bind people in related occupations. The question is the degree to which friendship and altruism among doctors, or among chief executives, is in their self interest rather than being only the result of their shared experiences. This in turn raises the question whether the manifestations of these intra-professional friendships, be they patient referrals or "Gary" dinners, are socially desirable.

7. References

Akerlof, George:"Loyalty Filters," *American Economic Review*, 73, March 1983, 54-63.

Aronson, Elliot: *The Social Animal*, New York: W.H. Freeman, 1984.

Becker, Gary S.: "A Theory of Social Interaction," *Journal of Political Economy*, 82, 1974a, 1063-91.

——: "Altruism, Egoism and Genetic Fitness: Economics and Sociobiology," *Journal of Economic Literature*, 14, September 1976.

——: *A Treatise on the Family*, Cambridge, Harvard University Press, 1991.

Becker, Gary S. and Kevin M. Murphy:"A Theory of Rational Addiction," *Journal of Political Economy*, 96, August 1988, 675-700.

Bergstrom, Theodore C.:"A Fresh Look at the Rotten Kid Theorem - and Other Household Mysteries," *Journal of Political Economy*, 97, October 1989, 1138-59.

Berkowitz, Leonard:"Group Standards, Cohesiveness and Productivity," *Human Relations*, 7, 1954, 509-19.

Bernheim, B. Douglas and Oded Stark:"Altruism within the Family Reconsidered: Do Nice

Guys Finish Last?," *American Economic Review*, 78, December 1988, 1034-45.

Bulow, Jeremy I., John D. Geanakoplos and Paul D. Klemperer:"Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of political Economy*, 93, June 1985, 488-511.

Camerer, Colin:"Gifts as Economic Signals and Social Symbols," *American Journal of Sociology*, 94, 1988 Supplement, S180-S214.

Coleman, James S.: *Foundations of Social Theory*, Cambridge, Harvard University Press, 1990.

Crozier, Michel: *The Bureaucratic Phenomenon*, Chicago, University of Chicago Press, 1964.

Dalton, Melville:*Men Who Manage*, New York, Wiley, 1959.

Farrell, Joseph and Garth Saloner: "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, 16, Spring 1985, 70-83.

Frank, Robert H.:"If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?," *American Economic Review*, 77, September 1987, 593-604.

Gladstein, Deborah:"A Model of Task Group Effectiveness," *Administrative Science Quarterly*, 29, 1984, 499-517.

Goodman, Paul S., Elizabeth Ravlin and Marshall Schminke:"Understanding Groups in Organizations," *Research in Organizational Behavior*, 1986.

Green, Jerry R. and Nancy L. Stokey:"A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, June 1983, 349-64.

Hirshleifer, Jack:"Shakespeare vs. Becker on Altruism: The Importance of Having the Last Word," *Journal of Economic Literature*, June 1977, 15, 500-502.

Holmström, Bengt:"Moral Hazard in Teams," *Bell Journal of Economics*, 13, Autumn 1982, 74-91.

—— and Paul Milgrom:"Regulating Trade among Agents", *Journal of Institutional and Theoretical Economics*, 146, March 1990.

Homans, George C.:*The Human Group*, New York, Harcourt, 1950.

Itoh, H:"Essays on the Internal Organization of the Firm," Unpublished Ph.d thesis, Stanford University, 1988.

Lazear, Edward P.: "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97, June 1989, 561-80.

—— and Sherwin Rosen: "Rank-Order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89, October 1981, 841-64.

Mayo, Elton: *The Human Problems of an Industrial Civilization*, New York, Macmillan, 1933.

Montgomery, James: "Revisiting *Tally's Corner*: Mainstream Norms, Cognitive Dissonance, and Underclass Behavior," mimeo, 1991.

Rabin, Matthew: "Incorporating Fairness into Game Theory", mimeo 1991.

Roethlisberger, F.J. and William Dickson: *Management and the Worker*, Cambridge, Harvard University Press, 1939.

Seashore, S.E.: *Group Cohesiveness in the Industrial Work Group*, Ann Arbor, Survey Research Center, 1954.

Spence, A. Michael: *Market Signaling*, Cambridge, Harvard University Press, 1974.

Stodgill, Ralph M.: "Group Productivity, Drive and Cohesiveness," *Organizational Behavior and Human Performance*, 8, 1972, 26-43.

Thibaut, John W. and Harold H. Kelley: *The Social Psychology of Groups*, New York, Wiley, 1959.

Tirole, Jean: "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organizations*, 2, Fall 1986, 181-214.

Date Due

Lib-26-67