TI 2009-027/1

Tinbergen Institute Discussion Paper

The Breakdown of Morale

Nick Vikander

Erasmus School of Economics, and Tinbergen Institute.

The Breakdown of Morale *

Nick Vikander †

February 2009

Abstract

This paper studies how morale in teams can break down. It interprets high morale as team members working together productively, either because of a sense of fairness or because of implicit incentives from repeated interactions. Team members learn that lay-offs will occur at a fixed future date, which will eventually cause morale to break down.

The paper shows that the breakdown of morale can vary in size and the equilibrium outcomes can be Pareto ranked. A firm's measures to encourage cooperation may actually hurt morale, by convincing opportunistic team members to imitate and later take advantage of cooperative colleagues.

1 Introduction

1.1 The Issue

Managers are preoccupied with employee morale. They feel it is important for productivity, but worry that high morale may be fragile. In particular, increased turnover or a sense of unfairness may cause morale to break down. Managers also worry that low morale can be contagious (Bewley 1999).

This paper examines a particular mechanism, based on these ideas, by which morale may break down. I look at team production, and define high morale as team members choosing to cooperate (work productively) rather than defect (be opportunistic).

Team members are of two types. Altruists are concerned with fairness and will cooperate if they expect others to do the same. Egoists only cooperate if repeated team interactions provide implicit incentives for them to do so. Team members have been working in a stationary environment where the incentives from repeated interactions have sustained high morale. Altruists have not distinguished themselves from egoists, so team members only know their own type.

*I would like to thank my supervisor Maarten Janssen, and seminar participants in Rotterdam, Amsterdam and Vienna for helpful comments.

†Erasmus University Rotterdam and Tinbergen Institute, vikander@tinbergen.nl, www.tinbergen.nl/∼vikander

I consider what happens if team members unexpectedly learn that a one-time shock, interpreted as lay-offs, will occur at a fixed future date. They more heavily discount periods after the shock, which is large enough to eventually cause egoist morale to break down. Low morale may spread from egoists to altruists, who do not want to be taken advantage of by egoist partners.

An equilibrium always exists where morale breaks down completely, so where players defect in all periods until the shock. Any cooperation is only possible if altruists can reveal their type by cooperating in some period where egoists defect.

A firm that makes cooperation sufficiently attractive can prevent any breakdown, but a smaller change may actually hurt morale. It encourages egoists to imitate altruists and later take advantage of them, which can leave altruists unable to reveal their type. As a result, all equilibrium cooperation can break down.

1.2 Literature on Morale, Teams

Morale is an ambiguous notion, related to trust in colleagues, a willingness to cooperate, a sense of common purpose and a belief that outcomes are fair. Little empirical work has looked at how morale in firms may break down, perhaps because firms are hesitant to make their problems public (Bewley 1999). But there is evidence supporting this paper's approach: to look at morale in self-managed teams, and at the impact of both concerns for fairness and incentives from repeated interactions.

Empirical work links high morale with high group performance, likely because it promotes cooperate behaviour (see for example Ryan et al. 1996). Cooperation is important in teams, since interdependent tasks and a shared responsibility for outcomes can otherwise make opportunism tempting (Alchian and Demsetz 1972). It may be particularly important in self-managed teams, where employers exert limited control.

Moreover, teams are now widely used in firms.[1] Self-managed teams have become particularly popular. Osterman (2000) reports that 40% of firms in the United States had at least half of their core employees in self-managed teams in 1997, and the practice is widespread internationally.[2]

Intrinsic motivation, such as concerns for fairness, and implicit incentives from repeated interactions can both explain why employees cooperate (Rotemberg 1994). Both reasons are compelling in the context of self-managed teams. Intrinsic motivation may increase as people come to believe in common values for their work based on fairness and identification with the team (Barker 1993). Many teams also interact repeatedly and team members' ability to discipline each other makes reputation important (Burt 2005, Che and Yoo 2001, Rayo 2007, Kvaloy and Olsen 2006).

[1] A government survey (Kinsley et al. 2005) reports that 72% of firms in Great Britain had at least some employees in formally designated teams in 2004.

[2] See Clegg et al. (2002) for Great Britain, Switzerland, Australia and Japan, and Wood et al. (2004) again for Great Britain.

This paper is related to the literature on team interactions and productivity. Holmstrom and Milgrom (1990) look at how binding contracts between team members influence their relationship with the principal, and Kandel and Lazear (1992) look at peer pressure between partners. Che and Yoo (2001) examine how repeated interactions affect a principal's choice of joint performance or relative performance evaluation, while Kvaloy and Olsen (2006) consider a similar situation with only relational contracts. Rayo (2007) looks at repeated interactions with cash transfers between team members.

This paper is different in that it looks at how incentives from repeated interactions interact with concerns for fairness, and relates the whole to the concept of morale. Among the above papers, it is also the only one to consider repeated interactions in a non-stationary environment. This type of environment is relevant since breakdowns in morale are often related to sudden changes in the state of the world (Gibbons 1998, Levin 2002). Another difference is that I assume all incentives come from repeated interactions between teams members in their tasks.

The strategic situation in this paper bears similarities to the finitely repeated prisoners' dilemma, which has been looked at extensively. Cooperation may be possible due to group exclusion (Hirshlifer and Rassmusen 1989), Knightian uncertainty (James and Costa 1994), dependence between past actions and current period pay-offs (Janssen et al. 1997), unilateral pre-game commitment (Faíña Medín et al. 1998), or rewarding pairs of players who outperform others (Serrano and Zapater 1998).

This paper's approach is most similar to Kreps et al. (1982), which considers the effect of uncertainty about player rationality or preferences. Kreps et al. look at mixed strategies, and show that the classic conclusion (no cooperation) is not robust to the introduction of a small probability each player is an altruist. This paper instead looks at symmetric pure strategy equilibria, and considers how a firm's policy decisions may affect the amount of cooperation.

1.3 What is Ahead

Section 2 describes the model, which involves an infinitely repeated game with two types and a shock at a fixed future date. The stage game is a prisoners' dilemma for egoists and a coordination game for altruists. I identify cooperation with high morale, and show to what extent it is possible before the shock.

Section 3 shows there are three types of equilibrium outcomes, and they can be Pareto ranked. The worst case is for morale to break down immediately upon learning of the shock and to remain low. An intermediate case is similar but altruist morale recovers at some point before the shock. The best case is for altruist morale to stay high throughout, and for egoist morale to only drop at some later time. It is only in this case that egoist morale does not immediately drop, and low morale never spreads to altruists.

I then describe what drives these equilibrium results. A key point is that cooperation is only possible if players eventually sort; altruists cooperate and egoists defect in some period, allowing altruists to reveal their type and work with one another. If not, then egoists eventually defect because of the shock, and altruists defect because of concern their partner is an egoist.

Seemingly reasonable steps a firm might take to help morale can be counterproductive. Increasing the returns to successful cooperation or targeting the lay-offs at those who defect can actually hurt morale. It decreases an egoist's incentive to sort, since he may prefer imitate an altruist and later defect against him. That can cause a sorting equilibrium to break down, leaving only the worst case outcome. On the other hand, giving players more flexibility about when they can act and informing them well in advance of the lay-offs can help morale.

Section 4 shows the general results are robust in terms of renegotiation proofness, different assumptions about rematching, and allowing for uncertainty about the shock. Section 5 concludes, and points out a parallel with other types of efficient groups that are unable to adjust to unexpected change.

2 The Model

2.1 Model Set-Up

There are a countably infinite number of players who differ only in type, $\theta \in \Theta$, where $\Theta = \{0, \theta_0\}$. Altruists have type $\theta = \theta_0$ and egoists have $\theta = 0$, with $\theta_0 > 0$. Type is private information, and a fraction λ of players are altruists.

Players play an infinitely repeated game, in which they pair up into a team in each period and collaborate on a productive task. Players learn at the start of the game that an exogenous shock will occur after T_0 periods, at which time a randomly selected group of players will leave the game. A fraction $\delta_0 < 1$ of players remain after period T_0, so a low value of δ_0 corresponds to a large shock. Players whose partner has left the game randomly match up with each other and play continues.

Players have a natural discount factor $\delta \leq 1$. They are uncertain to remain in the game after period T_0, so their effective discount factor looking forward to periods after the shock is $\delta\delta_0 < \delta$.

The following stage game gives the pay-offs for team production:

	C	D
C	$a+\theta, a+\theta$	c, b
D	b, c	d, d

The material incentives a, b, c and d represent a prisoners' dilemma: $b > a > d > c$, and $2a > b + c$.

Altruists are intrinsically motivated, receiving an extra pay-off $\theta_0 > b - a$ when their cooperation is reciprocated. Two altruists who knew each other's type would play a coordination game.

Players are randomly matched in the first period. Teams in which both members choose the same action in a given period remain together for the next period. All other teams break up, and these players are rematched into pairs according to some exogenous mechanism. The only assumption about the mechanism is that all players who are believed to be altruists with probability one are paired together.

I look for perfect Bayesian equilibria. Such equilibria consist of strategies and beliefs, such that strategies are always optimal given beliefs and beliefs follow from observed actions. Players have a common prior λ and they update their beliefs using Bayes' rule whenever possible.

2.2 Interpretation and Additional Assumptions

Players should be seen as employees working for a firm, with each pair working as a self-managed team that operates autonomously from any principal. Morale is high in a given period if players cooperate. Players can break up with their partner and rematch if they are unhappy with their partner, which often seems to be possible in self-managed teams (see for example Barker 1993).

The start of the formal game corresponds to players learning of the shock, at which point they do not know each other's type. The interpretation is that players were working together before then with high morale, supported by incentives from repeated interactions, so nobody could infer each other's type. In that sense, the first period where players defect is when morale breaks down. The initial matching is random because players have no reason to differentiate between each other based on past play.

The material incentives give a prisoners' dilemma because of the potential for opportunism in teams. It corresponds to a particular situation with joint production technology, group incentives and costly effort: productivity is $2d$, $2b$ or $2(b+a-c)$ when nobody, one player or two players exert effort, each player receives half his team's production as pay-off and the cost of effort is $(b-c)$.

The term θ_0 in an altruist's pay-off reflects a sense of fairness towards his partner. He is intrinsically motivated to cooperate, but only if he expects his partner to cooperate as well.

I only consider a subset of the potential perfect Bayesian equilibria. First, I only look at symmetric pure strategy equilibria, where all players of the same type take the same actions. Looking at symmetric pure strategy equilibria simplifies the analysis and allows me to consider the two types of employees as groups. I can then address how morale in one group affects morale in the other.

Second, I look for equilibrium with a certain amount of cooperation. Specifically, I only consider equilibrium outcomes in which players cooperate when they believe each other to be altruists with probability one, and in which all players cooperate after the shock. I do not place any restriction on play off the equilibrium path - in principle a player could punish a deviation by repeated defection. The reason is that I am interested in the maximum amount of cooperation that can be sustained in equilibrium, despite uncertainty about player type and the upcoming shock. I check the robustness of the results in Section 4 by applying renegotiation-proofness, which may restrict possible play off the equilibrium path.

I use the one-stage deviation principle when looking for potentially profitable deviations from any equilibrium. In any continuation game and for given beliefs, I only have to consider deviations in which a player changes his first stage action and then conforms to his equilibrium strategy (Fudenberg and Tirole 1991).

Finally, I make two assumptions on parameters. The discount factor must be large enough to support full cooperation if players did not expect a shock. Condition 1 is:

$$\delta \geq \frac{b-a}{b-d} \tag{1}$$

The shock must be large enough to cause an equilibrium in which players always cooperate to break down. Otherwise, morale could simply remain high. An egoist's best deviation is to defect in the last period before the shock. He obtains b in that period, and remains in the game with probability δ_0. If he remains he may be punished by repeated defection, which would give him d in all later periods. In the equilibrium, the egoist would receive a in the period before the shock, and also in all periods afterwards if he remains in the game. The deviation is profitable if:

$$a + \delta_0 \frac{\delta}{1-\delta} a < b + \delta_0 \frac{\delta}{1-\delta} d$$

Condition 2 is therefore:

$$\delta_0 < \frac{1-\delta}{\delta} \frac{b-a}{a-d} \tag{2}$$

Condition 1 implies that the upper bound on δ_0 is less than one.

3 Model Analysis

Theorem 1 will describe what outcomes can be supported in equilibrium. The outcomes can be Pareto ranked, and equilibria with higher morale only exist for more restrictive parameter values.

I establish a number of lemmas which I use to formulate and prove Theorem 1. The first lemma shows that altruists cooperate at least as much as egoists in any equilibrium.

Lemma 1. *In equilibrium, altruists cooperate in any period where egoists cooperate.*

Proof. Altruists and egoists fully reveal their type as soon as they choose different actions. After revealing their type, egoists defect until the shock by Condition 2 and by backwards induction. Before revealing their type, players choose the same actions so altruists cannot defect in a period where egoists cooperate. If players revealed their type by egoists cooperating and altruists defecting, an egoist could improve his pay-off by defecting. He would obtain $b > a$ in that period and at least the equilibrium pay-off d in each later period. □

An egoist's dominant strategy in the stage game is to defect. He will only cooperate if he can increase his future pay-off by imitating an altruist.

The second lemma shows that strategies of a simple form can generate all equilibrium outcomes. It is important to first distinguish between non-visible and visible deviations.

For a particular equilibrium, a player's deviation is non-visible if it fully mimics the equilibrium actions of the other type. It is non-visible in the sense that other players cannot infer that he is deviating. Rather, he appears to be the other type playing his equilibrium strategy. Any other deviation is visible.

Players can punish someone who makes a visible deviation, but have no reason to do so after a non-visible deviation. The lemma shows I can assume players always respond the same way after any visible deviation.

Lemma 2. *Consider any equilibrium outcome. It can be supported for arbitrary out-of-equilibrium beliefs by strategies where players always defect after a visible deviation.*

Proof. Any subgame perfect equilibrium outcome can be supported by equilibrium strategies that revert to the least attractive equilibrium play of the subgame following a deviation (Abreu 1988). Here the least attractive equilibrium play in any continuation game is repeated mutual defection. That is the case regardless of beliefs, since mutual defection is a Nash equilibrium of the stage game for both types.

□

I now only consider strategies where players always defect after any visible deviation. Doing so will be sufficient to find all possible equilibrium outcomes. To fully specify each type's strategy, I must also give actions on the equilibrium path for the first T_0 periods, and after making a non-visible deviation. With these strategies, a visible deviation can only be profitable if it provides an immediate gain compared to the equilibrium outcome.

I call an equilibrium where altruists always take the same actions as egoists a pooling equilibrium. Players could cooperate in all periods, defect in all periods, or cooperate in some periods and defect in others. The third lemma shows there is a unique pooling equilibrium.

Lemma 3. *The only pooling equilibrium is for all players to defect in periods* 1 *through* T_0.

Proof. Any deviation is visible since players take the same equilibrium actions, so the latter fully specify a strategy. It is an equilibrium, since cooperating in any period reduces immediate pay-off by $(d - c)$.

In any other pooling equilibrium, let $t < T_0$ be the last period before the shock where players cooperate. An egoist who defects in period t improves his immediate pay-off by $(b - a)$. Until the shock, he can do no worse than his equilibrium pay-off d. He only suffers a punishment if he remains in the game after the shock, receiving d instead of a in all later periods. The deviation is profitable if:

$$\delta^{t-1}(b-a) > \frac{\delta^{T_0}}{1-\delta}\delta_0(a-d)$$

The incentive to deviate is decreasing in t, and it is profitable for $t = T_0$ by Condition 2.

□

Lemma 3 implies any cooperation is only possible before the shock if players eventually sort. A sorting equilibrium is one where altruists take different actions from egoists in some period. I say players sort in a given period, if that is the first period where they reveal their type.

I look at sorting equilibria in the following way. I first assume players sort in a given period and describe how they then must play until the shock. I then show in what period players are willing to sort. Finally, I assume players sort in one such period and describe what play is possible before then.

Lemma 4. *Say players sort in period t. Then on the equilibrium path, altruists cooperate and egoists defect in period t and in all subsequent periods until the shock.*

Off the equilibrium path, an altruist who defects in period t and in all periods since then continues to defect until the shock. An egoist who cooperates in period t and in all periods since then continues to cooperate until period T_0, and then defects.

Proof. The lemma specifies a strategy for each type as of period t, because it gives actions on the equilibrium path and after any non-visible deviation.

Lemma 1 implies altruists cooperate and egoists defect in period t. Types are then fully revealed, so by the assumption on matching altruists pair up with each other. Altruists then cooperate until the shock because they know they are working with fellow altruists, and egoists defect by the same logic as in Lemma 1.

An altruist who defects in period t and all subsequent periods is believed to be an egoist, so his partner defects until the shock. Cooperating would decrease his pay-off by $(b-a)$ and make the deviation visible, so it is more profitable to keep on defecting.

An egoist who cooperates in period t and subsequent periods is believed to be an altruist, so his partner cooperates until the shock. He receives a in all periods, then b in period T_0 when he defects. That makes the deviation visible, so he gets d in each period after the shock if he remains in the game. Cooperating in period T_0 would give him a in all these periods, which is less profitable by Condition 2. First defecting before period T_0 is also less profitable by Condition 1.

□

Say players sort $T \equiv T_0 - t$ periods before the shock, so after t periods of play. Lemma 4 implies that when looking for a profitable deviation in these last

T periods, it is sufficient to consider a one-stage deviation in the period where players sort. That is, does either type have an incentive to change his action in that period, and conform to his equilibrium strategy thereafter?

If the one-stage deviation is not profitable for a player of a certain type, I say that type is willing to sort T periods before the shock. A sorting equilibrium can only exist if altruists and egoists are willing to sort in the same period.

The following two lemmas describe when altruists and egoists are willing to sort.

Lemma 5. *Altruists are willing to sort T periods before the shock if and only if $T_a \leq T$, where T_a solves:*

$$\lambda b + (1-\lambda)d - \lambda(a+\theta_0) - (1-\lambda)c = \frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0-d) \tag{3}$$

If no solution exists, then $T_a = \infty$. An altruist's incentive to sort is strictly increasing in T.

Proof. From Lemma 4, an altruist's equilibrium pay-off is:

$$\lambda(a+\theta_0) + (1-\lambda)c + \frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0) + \delta_0\frac{\delta^T}{1-\delta}(a+\theta_0)$$

A one-stage deviation by defecting when players sort gives:

$$\lambda b + (1-\lambda)d + \frac{\delta}{1-\delta}(1-\delta^{T-1})d + \delta_0\frac{\delta^T}{1-\delta}(a+\theta_0)$$

Comparing the two expressions gives (3), which is easier to satisfy for larger T. □

An altruist's equilibrium actions are to cooperate in all periods after sorting, while his best deviation is to imitate an egoist and defect in all these periods.

The left-hand side of (3) gives an altruist's immediate loss by sorting rather than deviating, and the right-hand side gives his future gains. These future gains are $(a+\theta_0-d)$ in all periods until the shock, so sorting is more attractive when T is large.

If sorting actually gives an immediate gain, then altruists are willing to sort for any T, so $T_a = 1$. That is the case when the left-hand side of (3) is negative. They are unwilling to sort for any T if sorting gives an immediate loss that is larger than even an infinite stream of future gains, so that $T_a = \infty$. Otherwise, there is a finite value of $T_a > 1$ such that altruists are only willing to sort if the shock is at least T periods away.

The term δ_0 does not appear in the altruist's incentive constraint because an altruist's best deviation is non-visible. He will not be punished after the shock, and will instead enjoy the equilibrium pay-off a in all these periods. An altruist's incentive to deviate therefore does not depend on the probability he will leave the game.

Lemma 6. *Egoists are willing to sort T periods before the shock if and only if $T \leq T_e$, where T_e is the maximum of 1 and the solution to:*

$$\frac{\delta}{1-\delta}(1-\delta^{T-1})(a-d)+\delta^{T-1}(b-a)-\delta_0\frac{\delta^T}{1-\delta}(a-d) = \lambda b+(1-\lambda)d-\lambda a-(1-\lambda)c \tag{4}$$

If no solution exists, then $T_e = \infty$. An egoist's incentive to sort is strictly decreasing in T.

Proof. Again, Lemma 4 gives players' equilibrium strategies as of T periods before the shock. An egoist will always sort when $T = 1$, because defecting is the dominant strategy of the stage game. If $T \geq 2$, an egoist gets pay-off:

$$\lambda b + (1-\lambda)d + \frac{\delta}{1-\delta}(1-\delta^{T-1})d + \delta_0\frac{\delta^T}{1-\delta}a$$

Cooperating when players sort gives:

$$\lambda a + (1-\lambda)c + \frac{\delta}{1-\delta}(1-\delta^{T-1})a + \delta^{T-1}(b-a) + \delta_0\frac{\delta^T}{1-\delta}d$$

Comparing the two expressions gives (4). It is more difficult to satisfy for larger T if the following expression is negative:

$$(b-a) - \frac{\delta}{1-\delta}(a-d)(1+\delta_0)$$

Plugging in $\delta_0 = 0$ and δ's lower bound from Condition 1 shows that is the case. □

An egoist's equilibrium actions are to defect in all periods after sorting, while his best deviation is to imitate an altruist but defect in period T_0.

The left-hand side of (4) gives an egoist's immediate gain by sorting rather than deviating, and the right-hand side gives his future losses when $T \geq 2$. The immediate gains are always positive since defecting is the dominant strategy of the stage game, and losses are positive and increasing in T. Sorting is therefore less attractive when T is large.

If the immediate gain from sorting is larger than an infinite stream of future losses, then egoists are willing to sort for any T (so $T_e = \infty$). Otherwise, there is a finite value of T_e such that egoists are only willing to sort if the shock is at most T_e periods away.

Lemmas 5 and 6 imply sorting can occur T periods before the shock if and only if:

$$T_a \leq T \leq T_e \tag{5}$$

The following lemma shows that T_e is always at least as large as T_a. When T_a is finite, a real number T will always exists that satisfies (5).

Lemma 7. *$T_a \leq T_e$ holds for all parameter values.*

Proof. See appendix □

Players must pool in all periods before they sort, and Lemma 8 shows if it is possible to pool on cooperation.

Lemma 8. *Say players pool for t periods and then sort. Then no player has an incentive to deviate before sorting if players always pool on defection, or if the last period $t' \leq t$ where players pool on cooperation satisfies:*

$$\lambda \geq \frac{1}{\delta^{t-t'+1}(b-d)}[(b-a) - \delta_0 \frac{\delta^{T_0-t'+1}}{1-\delta}(a-d)] \tag{6}$$

Proof. Any deviation in the first t periods is observable since players are pooling. The only deviation that increases immediate pay-off is for an egoist to defect while all other players cooperate. Condition 1 implies the deviation is most attractive in period t'. His equilibrium pay-off from then on is:

$$\delta^{t'-1}a + \frac{\delta^{t'}}{1-\delta}(1-\delta^{t'-t})d + \delta^t\{\lambda b + (1-\lambda)d\} + \frac{\delta^{t+1}}{1-\delta}(1-\delta^{T_0-t})d + \delta_0\frac{\delta^{T_0}}{1-\delta}a$$

He gets a in period t', d in the following periods until sorting, d after sorting and until the shock, and then a in subsequent periods if he remains in the game.

His best deviation is to defect in period t' and in all subsequent periods, giving:

$$\delta^{t'-1}b + \frac{\delta^{t'}}{1-\delta}(1-\delta^{T_0-t'+1})d + \delta_0\frac{\delta^{T_0}}{1-\delta}d$$

Comparing the two expressions gives (6).

□

Lemma 8 shows that players can always defect before sorting. Cooperation is only possible before sorting if the fraction of altruists exceeds a certain threshold.

If other players cooperate before sorting and λ is low, an egoist prefers to defect in the last period where players cooperate. He is then punished and receives only d instead of $[\lambda b + (1-\lambda)d]$ when sorting occurs, but these pay-offs are similar when λ is low. Plugging the upper bound of δ_0 into (6) shows that the threshold is always positive.

The threshold depends on t', the last period where players cooperate. It is decreasing in t', and so easiest to satisfy later in the game. Setting $t' = t$ in (6) and using $T \equiv T_0 - t$, the condition for cooperation in all periods before sorting is:

$$\lambda \geq \frac{1}{\delta(b-d)}[(b-a) - \delta_0\frac{\delta^{T+1}}{1-\delta}(a-d)] \tag{7}$$

where the right-hand-side is less than one by Condition 1.

If it is possible to cooperate at all before sorting, then it is possible to cooperate in all periods before sorting. This later case is most reasonable, as it is Pareto dominant and more intuitive than players switching between cooperation and defection.

Combining the results from the lemmas gives Theorem 1, which shows what outcomes can be supported by equilibrium strategies. All actions refers to the first T_0 periods of the game.

Theorem 1. *An equilibrium always exists where players defect in all periods. That is the only equilibrium if:*

$$T_0 < T_a$$

or if there is no integer T such that (5) holds:

$$T_a \leq T \leq T_e$$

where T_a and T_e are defined in Lemmas 5 and 6.

If instead $T_0 \geq T_a$ and (5) holds for some T, then a second type of equilibrium exists. Egoists defect in all periods, while altruists only cooperate in the last T periods before the shock.

If λ is large enough so that (6) also holds, then a third type of equilibrium exists. Players pool on either cooperation or defection in any of the first $t' - 1 < T_0 - T - 1$ periods, and they cooperate in period t'. Egoists defect in all later periods, while altruists only cooperate in the last T periods before the shock.

There are no other equilibria.

Proof. The proof follows from Lemmas 1-8.

□

I now consider only the most intuitive equilibrium of the third type, where players always cooperate before sorting. I interpret the results in terms of morale, and then describe what drives them.

The equilibrium outcomes can be Pareto ranked. The worst case for morale is where all players defect until after the shock. Egoist morale drops upon learning of the shock and low morale spreads immediately to altruists. Morale then remains low until the shock. This equilibrium exists for all parameter values.

An intermediate case is similar but where altruists return to cooperation some time before the shock. Egoist morale drops immediately and low morale spreads to altruists, but altruist morale recovers before the shock. These equilibria exist for certain parameter values.

The best case for morale is where altruists always cooperate, and egoists initially cooperate but later begin to defect. Low egoist morale never spreads to altruists, whose morale remains high throughout. These equilibria exist for more restrictive parameter values.

If type were public information, altruist morale would always remain high but egoist morale would drop immediately. Welfare would be higher than in the first two cases but lower than in the third.

3.1 What Drives the Equilibrium Results

Altruists can only sustain high morale if they know they will eventually be able to identify one another. To do so, they must sort by cooperating in some period where egoists defect, which reveals their type. Sorting may involve an immediate loss compared to deviating, but yields future gains since altruists can then cooperate with each other.

All morale will break down by backwards induction if altruists are never able to sort. Altruists know egoist morale will be low in the last period before the shock. Their own morale will also be low if they won't sort, which ensures egoist morale will also be low the period before. If altruists won't sort then either, egoist morale will drop even earlier. Working back implies egoist morale drops in the first period, which causes altruist morale to drop as well.

If altruists are eventually able to identify each other, what happens to morale depends on the fraction of altruists. As long as altruists cooperate, egoists may be convinced to do the same to avoid revealing their type. Even then, it may be profitable for an egoist to defect in the last period where all players cooperate. He is punished when players sort, but the deviation is attractive if his partner was most likely an egoist who would defect then anyway. In a sense, egoist morale remains low because there are too few altruists to change their behaviour.

If egoist morale immediately drops, altruists would like to sort by cooperating and keep their morale high. But cooperating when $T > T_e$ is so attractive that some egoists would imitate and later take advantage of them. Altruist morale therefore drops and remains low until $T \leq T_e$, then recovers as players sort.

If there are enough altruists and they are eventually able to identify each other, altruist morale can remain high the entire game. High altruist morale keeps egoist morale high until players sort. Egoist morale then drops, and altruists continue cooperating with one another.

3.2 Allowing For Flexbility

Morale may completely break down because players are unwilling to sort at the start of any period, despite the fact that $T_a < T_e$. If there is no integer between T_a and T_e, then any given period is either too early or too late to sort.

This section shows a complete breakdown can be avoided if players have enough flexibility, in the sense of choosing actions more frequently. Intuition suggests that if periods are sufficiently short, players will be able to sort between time T_a and T_e before the shock. That intuition does not turn out to be fully correct, but an equilibrium will nonetheless exist where players cooperate.

To model flexibility, I scale down per period pay-offs by an integer $M \geq 1$ and assume there are MT_0 periods before the shock. Each period now has length $\frac{1}{M}$, but the total time until the shock is unchanged. I assume actions are revealed to other players time τ after they occur, with $0 < \tau \leq 1$. It reduces to the original model for $M = 1$, since then players can update their beliefs and act on them after each period.

For technical reasons, I only consider the case of δ close to 1. Intuitively, the weight of pay-offs in one time interval compared to another should not depend on M. For example, the total pay-off before the shock compared to the total pay-off afterward should be independent of period length. Experimenting shows this is not generally possible for geometric discounting, so I set $\delta = 1$ when needed in the formal analysis. By continuity, the results will also hold for δ close to one.

I first show that if τ is sufficiently small, increasing flexibility can stop morale from breaking down at all.

Result 1. *Say $\tau \leq \delta_0 \frac{\delta}{1-\delta} \frac{(a-d)}{(b-a)}$, and $\delta_0 > 0$. Then for M sufficiently large, an equilibrium exists where players always cooperate.*

Proof. Say $\tau \leq \frac{1}{M}$, so actions are revealed after each period. In that case, if all players cooperate, an egoist does not have an incentive to defect in the last period before the shock if:

$$\frac{b-a}{M} - \delta_0 \frac{\delta}{1-\delta}(a-d) \leq 0$$

That reduces to $\frac{1}{M} \leq \delta_0 \frac{\delta}{1-\delta} \frac{a-d}{b-a}$, which can hold if $\tau \leq \delta_0 \frac{\delta}{1-\delta} \frac{a-d}{b-a}$ □

As long as each period is longer than τ, increasing flexibility reduces an egoist's incentive to defect in the last period before the shock. He can then only take advantage of his partner for a shorter time before his defection is discovered. If some players remain in the game after the shock and τ is small enough, then sufficiently flexibility makes egoists willing to cooperate.

If each period is shorter than τ, increasing flexibility does not change this incentive. Regardless of period length, he can only take advantage of his partner for time τ.

If τ is too small to permit full cooperation, then increasing flexibility can still help players sort. They will not do so between time T_a and T_e before the shock, but only later in the game.

Result 2. *Say $T > T_a$ and $T_e \geq 2$, but there is no integer between T_a and T_e. Then for all M such that $\tau < \frac{1}{M}$, no sorting equilibrium exists. For M sufficiently large, an equilibrium exists in which players sort between time τT_a and τT_e before the shock.*

Proof. See appendix

□

Increasing flexibility does not help players sort as long as each period is longer than τ. Players can choose actions more often, but egoists are only willing to sort later in the game because they can take advantage of their partner for a shorter time. The proof shows that sorting is only possible between time $\frac{T_a}{M}$ and $\frac{Te}{M}$ before the shock. As M increases, the interval begins at a later point in the game and it also becomes smaller. If there is no integer T with $T_a \leq T \leq T_e$, then players are still unwilling to sort in any period.

If each period is shorter than τ, actions are always revealed approximately time τ after they occur. Increasing M does not change this, so players are still willing to sort between time $T_a\tau$ and $T_e\tau$ before the shock. For sufficiently high M, some period will begin within this interval.

3.3 Policy - Reducing the Size of the Breakdown

Interpreting the shock as lay-offs, what can the firm do to favour high morale? A sufficiently large change in material pay-offs that makes cooperation more attractive can prevent any breakdown from occurring at all. That is, making $(b-a)$ small enough compared to $(a-d)$ will violate Condition 2, so that full cooperation becomes an equilibrium.

If such a large change is not possible, the firm can try to minimize the size of the breakdown. To do so it wants player to sort, which implies both $T_a \leq T$ and $T \leq T_e$ in some period, with T_a and T_e defined in (3) and (4). It also wants players to cooperate before sorting, for which (7) must hold. A change in parameters will only have an unambiguous positive effect on morale if it helps with at least one incentive constraint without hurting any of the others. That is, it decreases T_a, increases T_e, or loosens (7).

Changing the material rewards a, b, c, or d tends to have an ambiguous effect on morale. Sorting involves cooperation from altruists and defection from egoists, so increasing the incentive to sort for one type tends to decrease the incentive to sort for the other. Looking at (3) and (4), there is no change to any of these parameters that always helps morale, or that always hurts.

In particular, two seemingly reasonable steps by a firm to help morale can be counterproductive. The first is a small increase in a, the returns to successful cooperation. The second is to make a rule that links lay-offs to performance, such that players who defect are more likely to be laid-off. Both changes make cooperation more attractive, but they can cause all equilibrium cooperation before the shock to break down. The reason is that they may prevent sorting.

Theorem 2. *Say θ_0 is small. Then pay-off parameters a, b, c and d exist such that there is an equilibrium with some cooperation, but either of the following changes leaves all players defecting as the unique equilibrium:*

(i) there is a marginal increase in a

(ii) δ_0 is sufficiently small, and a player only leaves the game after the shock if all players who have defected in strictly more periods than him also leave the game.

Proof. See appendix □

Both of these changes are too small to convince egoists to always cooperate. The condition on δ_0 in (ii) means an egoist will likely leave the game regardless of his actions, so the rule increases his incentive to cooperate only slightly.

Equilibrium cooperation is still only possible if employees sort, and egoists are now less willing to do so.

The idea in the proof is to choose parameters such that egoists are indifferent about sorting in some period. When θ_0 is small, egoists and altruists have similar pay-offs so the critical values T_a and T_e are close to one another. Sorting is only possible in the period where egoists are indifferent, and each change gives them a strictly positive incentive to deviate.

The parameters used to prove Theorem 2 are not unique: d is unspecified, and c may take on any one of $(T_0 - 1)$ possible values. I can also replace each parameter by another one within a small neighbourhood, and the results will still hold for a slightly larger increase in a. The proof considers the standard model with period length 1, but the argument also holds for any strictly positive period length.

Theorem 2 bears some similarities to results from the literature on intrinsic motivation and crowding out. Gneezy and Rustichini (2000) show large monetary rewards often improve behaviour, but small rewards may actually hurt. One explanation is that offering money can ruin the signaling effect of a certain action, by encouraging people who are just motivated by monetary rewards (Janssen and Mendys 2004). Although the mechanism in this paper is different, the problem is still that a small increases in material compensation can encourage lower types to take a positive action (cooperating), which can prevent higher types from taking that action in equilibrium.

Rather than changing pay-off parameters, the firm can help morale by giving employees more flexibility to react to the shock. Increased flexibility, in the sense of choosing actions more frequently, can help in two different ways. If employees can observe each other's actions quickly enough, then flexibility means they can only be taken advantage of for a short time. Morale can then stay high because egoists never defect. If not, then increasing flexibility can still help morale by allowing sorting. When periods are sufficiently short, employees are able to sort right at the moment when both types are willing to do so.

The firm may also prevent a complete breakdown of morale by informing employees of the lay-offs well in advance. If $T_0 < T_a$, then morale breaks down because there is not enough potential future cooperation to convince altruists to reveal their type. That need not be the case if the firm informs employees at least time T_a before lay-offs occur. Still, informing employees earlier may be risky. If there is no integer between T_a and T_e, then employees will still be unwilling to sort. Morale will break down immediately, and now it will remain low for longer.

The size of the shock and the fraction of altruists are not factors the firm can easily control, but they do both affect morale. A smaller shock is better for morale, but only because it affects egoist behaviour. An egoist is more willing to sort, because he will likely to remain with the firm and can be punished for a deviation. In contrast, the size of the shock has no effect on an altruist's incentive to sort. An altruist's best deviation involves perfectly mimicking egoist equilibrium actions. He not punished because nobody realizes he has actually deviated.

A larger fraction of altruists may actually have a negative effect on morale. More altruists means that more employees cooperate during sorting. That increases an altruist's incentive to sort, since he is less likely to be taken advantage of. If $b - d > a - c$, then it also increases an egoist's incentive to sort. The condition means there are no strategic complementarities in terms of material pay-offs. Egoists are more willing to defect when they think their partner will cooperate, which favours sorting. But if there are such strategic complementarities, an increase in the fraction of altruists increases an egoist's incentive to cooperate, which may leave him unwilling to sort.

4 Robustness and Extensions

4.1 Renegotiation-Proofness

The results of Theorem 1 are general in the sense of holding for arbitrary out-of-equilibrium beliefs. The strategies involve punishments that are always formally credible, but they do not always seem equally reasonable. Two players off the equilibrium path who believe each other to be altruists with a high probability are supposed to defect, but they might well renegotiate to mutual cooperation.

I address the point by showing that a renegotiation-proof equilibrium exists for some out-of-equilibrium beliefs, provided a condition on parameter values holds. In this equilibrium, players play as efficiently as possible after every history given that they must do so in future periods as well.

The renegotiation-proof equilibrium outcome will be one of the equilibrium outcomes from Theorem 1, where the particular outcome depends on parameter values. Theorem 1 shows that different equilibria exist for different parameter values, but outcomes of existing equilibria can always be Pareto ranked. For any given parameters, the renegotiation-proof equilibrium will give the Pareto dominant outcome of the existing equilibria.

I am interested in play before the shock, so I apply the standard notion of renegotiation-proofness in a finitely repeated game, Pareto perfection (Bernheim et al. 1987), to play in the first T_0 periods. In this context, an equilibrium is renegotiation-proof if it satisfies the following criteria after any history h_t. Players play an equilibrium of the stage game in period T_0 that is Pareto efficient, taking into account the period $T_0 + 1$ continuation pay-off. They choose actions in period $T_0 - 1$ that are supported by the period T_0 continuation pay-offs, taking into account possible restrictions on period T_0 play. The period $T_0 - 1$ actions give a Pareto efficient outcome to the last two periods of play, given the possible restrictions. Players choose actions in period $T_0 - 2$ that are supported by the continuation pay-offs in period $T_0 - 1$ and that are efficient as above, and so on back to period t.

Let h_t be an arbitrary history, with T periods left until the shock. The history h_t implies a certain belief μ_i about each player $i = 1, 2, 3 \ldots$. Beliefs about different players may differ, and they need not correspond to the prior λ.

Consider the outcomes that can be supported as perfect Bayesian equilibria in the continuation game following h_t, by pure symmetric strategies. Here symmetric means players of the same type and facing an identical strategic situation take the same action. There is at least one such outcome, where any team i, j with $\mu_i = \mu_j = 1$ cooperates and any other team defects.

Define RPS, a renegotiation proof strategy, in the following way. When restricted to the continuation game following h_t, RPS picks out one such outcome that is Pareto efficient, and this for any h_t.

Lemma 9. *Say there are T periods left before the shock. RPS will have any player i and his partner defect regardless of type if $\mu_i < \mu_a^T$, where:*

$$\mu_a^T \equiv \frac{d - c - \frac{\delta}{1-\delta}(1 - \delta^{T-1})(a + \theta_0 - d)}{a + \theta_0 + d - b - c}$$

Proof. See appendix

□

In the general game, an altruist is unwilling to sort T periods before the shock if $T < T_a$. The only possible play is then repeated defection. The condition $\mu_i < \mu_a^T$ is equivalent to $T < T_a$, provided beliefs μ_i replace the prior λ in the definition of T_a. Lemma 9 implies a player is unwilling to cooperate if he believes it sufficiently likely his partner is an egoist. The critical value μ_a^T is decreasing in T, so if $\mu_i < \mu_a^T$ holds for some T, then RPS will also have player i and his partner defect in all later periods.

Beliefs must satisfy $\mu_i \geq 0$, so $\mu_i < \mu_a^T$ is only possible in all periods if:

$$d - c > \frac{\delta}{1-\delta}(1 - \delta^{T_0 - 1})(a + \theta_0 - d) \tag{8}$$

Otherwise a player may cooperate even if he is sure his partner is an egoist, because the future benefits of rematching with an altruist are so large. The condition is more likely to hold when discounting is large and when T_0 is small.

Since RPS implies the Pareto efficient perfect Bayesian equilibrium of any subgame, it also implies the Pareto efficient equilibrium for the entire game. I now show conditions under which RPS is an equilibrium strategy.

Theorem 3. *Say (8) holds, and out-of-equilibrium beliefs about a player who deviates T periods before the shock satisfy $\mu < \mu_a^{T-1}$. Then* RPS *supports the Pareto efficient equilibrium of the game as a renegotiation-proof equilibrium.*

Proof. By (8), the critical value for μ is strictly positive in any period. By Lemma 9 and the assumption on beliefs, a player who makes an observed deviation must then repeatedly defect until the shock. That is the harshest possible punishment, so RPS restricted to any continuation game is an equilibrium strategy.

Renegotiation-proofness does not restrict the possible punishments with these out-of-equilibrium beliefs. It is therefore equivalent to playing the Pareto efficient equilibrium in each continuation game, which is what RPS does.

□

The renegotiation-proof-equilibrium described above gives one of three outcomes depending on parameter values. The first is for morale to immediately drop for all players and remain low until after the shock. The second looks the same except altruist morale recovers before the shock in the first period where sorting is possible. The third is for altruist morale to always remain high, and egoist morale to first stay high then drop in the last period where sorting is possible.

4.2 Not Allowing for Rematching

The qualitative results from Section 3 still hold if each player must remain with his initial partner throughout the game. By an identical argument, cooperation is only possible before the shock if players can sort.

An altruist now has a lower incentive to sort, because he cannot rematch if his partner turns out to be an egoist. That reduces the expected future benefits from sorting by a factor λ. He is willing to sort T periods before the shock if:

$$\lambda b + (1-\lambda)d - \lambda(a+\theta_0) - (1-\lambda)c \leq \lambda[\frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0-d)] \quad (9)$$

An egoist now has a higher incentive to sort because his best deviation is less attractive. Imitating an altruist only yields benefits if his partner turns out to be an altruist. He is willing to sort if:

$$\lambda[\frac{\delta}{1-\delta}(1-\delta^{T-1})(a-d)+\delta^{T-1}(b-a)-\delta_0\frac{\delta^T}{1-\delta}(a-d)] \leq \lambda b+(1-\lambda)d-\lambda a-(1-\lambda)c \quad (10)$$

The only difference between conditions (9) and (10) and the conditions with rematching, (3) and (4), is the factor λ on one side of each inequality.

The condition for sorting T periods before the shock remains of the form $T_a \leq T \leq T_e$, but both T_a and T_e are now larger. The condition for cooperation before sorting, (6), remains the same. It comes from an egoist's incentive to defect in the last period of cooperation before sorting, which is not affected by the assumption on rematching.

The qualitative results from Theorem 1 therefore still hold, although the periods where sorting is possible may differ. A sorting equilibrium will also look somewhat different, since only altruists who are matched together can cooperate after sorting. An altruist who discovers he is matched with an egoist will instead have to defect until the shock.

Differentiating both sides of (9) and (10) shows how changes in parameter values affect both types' incentive to sort. Just as before, there are no changes in parameters that always have an unambiguously positive or negative effect on morale.

The results are now more robust in terms of renegotiation proofness. Rearranging (9) and substituting μ for λ gives the condition:

$$\mu \geq \frac{d-c}{a+\theta_0+d-b-c+\frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0-d)}$$

The right-hand side is always positive, since an altruist who knows he is matched with an egoist will never cooperate. By the same argument as in Section 4.1, a renegotiation-proof equilibrium exists if out-of-equilibrium beliefs satisfy the above inequality. Therefore for any parameter values, there are out-of-equilibrium beliefs such that a renegotiation-proof equilibrium exists.

4.3 Uncertainty About the Shock

Instead of knowing lay-offs will occur at a fixed date, employees may believe there is a future period where lay-offs are more likely. There might be an upcoming announcement on firm performance or a discussion of restructuring, and the uncertainty could also last some time. This section shows that such a belief will still cause morale to break down, if lay-offs are sufficiently likely.

Assume a shock can occur after any one of periods $T_0, \ldots, T_0+N-1$ for $N \geq 1$. The shock occurs with probability $p \leq 1$ after any such period, conditional on in not having occurred before. The set-up reduces to the original model for $N = p = 1$.

For morale to break down and the previous results to hold, I must show an egoist will defect if other players always cooperate. For $N = 1$ and $p < 1$, the condition in period T_0 is:

$$a + [p\delta_0 + (1-p)]\frac{\delta}{1-\delta}a < b + [p\delta_0 + (1-p)]\frac{\delta}{1-\delta}d$$

since an egoist remains in the game with probability $p\delta_0 + (1-p)$. That is:

$$p > \frac{1}{1-\delta_0}[1 - \frac{1-\delta}{\delta}\frac{b-a}{a-d}] \equiv p^1 \tag{11}$$

Condition 2 implies $p^1 < 1$, so an egoist will defect in period T_0 if he believes the shock to be sufficiently likely.

I now show that a critical value p^N exists for any N.

Lemma 10. *Say players cooperate in all periods. Then a critical value p^N exists, such that an egoist has an incentive to defect in period T_0 if and only if $p > p^N$.*

Proof. See appendix □

An egoist's expected punishment increases if he expects to be in the game in future periods. That probability is decreasing in p, which implies the existence of a critical value p^N.

I now show p^N is decreasing in N. That is, the more periods there are in which the shock might occur, the less likely it needs to be in any one period to cause morale to break down.

Note first that defecting in period T_0 gives an expected future loss of $(a-d)$ per period, taking into account discounting and the probability the egoist will remain in the game in that period. The future loss is $(a-d)$ times:

$$\sum_{i=1}^{N-1}[\delta(1-p)]^i + \frac{\delta^N}{1-\delta}(1-p)^N + \frac{\delta}{1-\delta}p\delta_0\sum_{i=0}^{N-1}[(1-p)\delta]^i \tag{12}$$

The first summation refers to each period $T_0+1,\ldots,T_0+N-2$, if the shock has not previously occurred. The second term refers to period T_0+N-1 and all later periods if the shock never occurs. The last summation refers to a period in which the shock occurs and all later periods.

Lemma 11. *The critical value p^N is decreasing in N.*

Proof. An egoist's immediate gain from deviating always equals $(b-a)$, so it is sufficient to show that (12) is decreasing in N. Taking (12) evaluated at $N+1$ and subtracting it from (12) gives:

$$\delta^N(1-p)^N + \frac{\delta^{N+1}}{1-\delta}(1-p)^{N+1} + \frac{\delta^{N+1}}{1-\delta}(1-p)^N p\delta_0 - \frac{\delta^N}{1-\delta}(1-p)^N$$

Simplifying gives the following, which is negative since $\delta_0 < 1$:

$$\frac{\delta}{1-\delta}[\delta(1-p)]^N(\delta_0 - 1)$$

□

A large value of N means the egoist is more likely to eventually leave the game, which makes the deviation more attractive. Lemma 11 also implies that for a given N, an egoist's incentive to defect is higher in period T_0 than at any later moment. If he cooperates in period T_0 and the shock does not occur, he faces an equivalent situation in period T_0+1 but with N reduced by one.

The critical value p^N is actually decreasing with N at a decreasing rate, and I now calculate the limit as N tends to infinity. That gives a lower bound on the value of p that can still cause morale to break down. That is also useful because it may be difficult or even impossible to derive an explicit expression for p^N.

Lemma 12. *As N tends to infinity, the critical value p^N tends to:*

$$p^\infty = \frac{(1-\delta)([\frac{1-\delta}{\delta}\frac{b-a}{a-d}])}{1-\delta_0-\delta(1-[\frac{1-\delta}{\delta}\frac{b-a}{a-d}])}$$

Proof. Letting N go to infinity in (12) gives:

$$\frac{\delta(1-p)}{1-\delta(1-p)} + \frac{\delta}{1-\delta}p\delta_0\frac{1}{1-\delta(1-p)}$$

$$\frac{\delta}{1-\delta}\frac{1}{1-\delta(1-p)}[(1-p)(1-\delta)+p\delta_0]$$

The following condition implies an egoist defects in period T_0, and rearranging completes the proof:

$$\frac{\delta}{1-\delta}\frac{1}{1-\delta(1-p)}[(1-p)(1-\delta)+p\delta_0](a-d)<(b-a)$$

□

To illustrate the effect of increasing N, say $b=4, a=3, d=2, \delta=\frac{2}{3}$ and $\delta_0=\frac{1}{3}$. Then $p^1=\frac{3}{4}$ and $p^\infty=\frac{1}{2}$. The critical value decreases at the fastest rate for small N, so N need not be extremely large to see a comparatively large drop in the critical value. With the above parameters, $p^2=0.59$.

5 Conclusion

High morale, interpreted as cooperation (productive work) within the team, can break down if players foresee that an upcoming shock will create turnover. The paper looks at shocks that are sufficiently large to always cause some type of drop in morale. Fewer expected future interactions eventually cause egoist morale to break down, which may in turn trigger a drop in altruist morale. Altruists will cooperate if others do, but they may begin to defect because of concern they are working with egoists.

Morale can break down to different extents, and the corresponding equilibria can be Pareto ranked. The worst case is a complete collapse upon learning of the shock, which occurs if players cannot sort and reveal their type. An intermediate case is similar but altruist morale eventually recovers before the shock. It is only in the best case, where players can sort and there are enough altruists, that egoist morale remains high for some time and low morale never spreads to altruists.

Interpreting the shock as lay-offs, I look at steps the firm can take to minimize the breakdown of morale. Seemingly reasonable steps such as targeting the lay-offs at players who defect or increasing the returns to successful cooperation can actually hurt morale. They can do so by encouraging egoists to imitate altruists when they cooperate to later take advantage of them, which can prevent altruists from revealing their type. A sufficiently large change favouring cooperation will prevent morale from breaking down at all, but small changes may be counterproductive.

Flexibility, in the sense of letting players choose actions more frequently, can help morale in two ways. It can reduce the time an egoist takes advantage of a cooperating partner, which may prevent morale from breaking down at all. It can also permit sorting by making it possible to act at the precise moment both types are willing to sort.

The firm may also prevent a complete breakdown of morale by informing employees of the lay-offs well in advance. That ensures there is enough value in future cooperation for an altruist to reveal his type, which favours sorting. However, informing employees earlier can be risky. If a sorting equilibrium still does not exist, then morale will still break down and will remain low for longer.

Large lay-offs are particularly bad for morale, but only because they influence egoist behaviour. Egoists are more less likely to remains with the firm, and therefore to be punished for any deviation. A larger fraction of altruists can actually hurt morale if there are strategic complementarities to material pay-offs, so if $b - d < a - c$. An egoist's partner is more likely to cooperate, which reduces his incentive to defect and sort.

The results are robust in terms of renegotiation proofness, different assumptions about rematching, and allowing for uncertainty about the shock.

Low morale could never spread to altruists if they knew each other's type at the beginning of the game. The interpretation is that they don't know each other's type because morale was always high before players learned of the shock. All players cooperated, which was good for efficiency but also means players never sorted. Altruists remained cut off from private information about each other, information which could have helped them adjust to the unexpected news about the shock.

The point is more general. The very characteristic that makes groups efficient can cut them off from information, leaving then unable to adjust to an unexpected change. Closed groups within organizations can promote trust and cooperation but may remain isolated from innovative ideas (Burt 2005). Organizations may develop specific information channels and improve efficiency, but neglect others and become unresponsive to changes in the environment (Arrow 1974). They may also become dependant on a cooperative network partner and lack a safety net if their partner unexpectedly exits (Uzzi 1996). This paper shows members of efficient groups may not only be cut off from important information about the outside world, but also from important information about each other.

A firm can inform employees about each other if it can make a clear distinction between altruists and egoists while times are still good. Egoists may be willing to accept such a distinction at a low cost, because it will only harm them during bad times which they do not foresee. Microsoft tried to do just that by creating a clearly defined group of permatemps (Levin 2002). The permatemps often worked alongside regular employees, but they could be easily let go if times got tough. The regular employees could then continue with their work without suffering a drop in morale.

Appendix

Proof of Lemma 7. Altruists are willing to sort when T satisfies (3):

$$\lambda b + (1-\lambda)d - \lambda(a+\theta_0) - (1-\lambda)c) \leq \frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0-d)$$

Egoists are always willing to sort when $T = 1$, and for larger T that satisfies (4):

$$\frac{\delta}{1-\delta}(1-\delta^{T-1})(a-d)+\delta^{T-1}(b-a)-\delta_0\frac{\delta^T}{1-\delta}(a-d) \leq \lambda b+(1-\lambda)d-\lambda a-(1-\lambda)c$$

The result holds for $T_a = 1$ since by definition $T_e \geq 1$. If T_a is infinite so that (3) does not hold for large T, then inspection shows that (4) cannot hold either. That means T_e is also infinite.

Say T_a and T_e are finite, with $T_a \geq 2$. Let T take on the real value such that the egoist is indifferent about sorting, so where (4) holds with equality. Then (3) holds if:

$$\frac{\delta}{1-\delta}(1-\delta^{T-1})(a-d)+\delta^{T-1}(b-a)-\delta_0\frac{\delta^{T_0}}{1-\delta}(a-d) < \frac{\delta}{1-\delta}(1-\delta^{T-1})(a+\theta_0-d)+\lambda\theta_0$$

It is sufficient to show the inequality holds for the case $\theta_0 = b-a$ and $\delta_0 = 0$. In that case it reduces to:

$$[\delta^{T-1}-\frac{\delta}{1-\delta}(1-\delta^{T-1})](b-a) < \lambda\theta_0$$

which holds because the left-hand side is negative.

Proof of Theorem 2. Let $T_0 \geq 2$, $N \geq 1$, $\delta = 1$ and $\delta_0 = 0$. The same argument would hold for $\delta < 1$ and $\delta_0 > 0$ Choose integer $T \leq T_0$ such that the egoist's incentive constraint is satisfied with equality with T periods remaining.

$$T_e = 1+(1-\lambda)\frac{(a+d-b-c)}{a-d} = T$$

One way to do so is to specify the value of c for any such T:

$$c = a+d-b-\frac{1}{1-\lambda}(a-d)(T-1)$$

This is consistent with $a+d > b+c$, since the last term on the right-hand-side is strictly positive.

Let ϵ be small and strictly positive, and let $\theta_0 = 2\epsilon$, $a = b-\epsilon$. Then:

$$T_e - T_a = (1-\lambda)\frac{a+d-b-c}{a-d} - \frac{(1-\lambda)(a+\theta_0+d-b-c)-(a+\theta_0-b)}{a+\theta_0-d}$$

$$T_e - T_a = (1-\lambda)\frac{d-c-\epsilon}{a-d} - \frac{(1-\lambda)(d-c+\epsilon)-\epsilon}{a+2\epsilon-d}$$

Bringing the terms over a common denominator gives an expression proportional to ϵ. For given N, choose ϵ such that $T_e - T_a < \frac{1}{N}$.

Players are only willing to sort T periods before the shock, since $T_e = T$ implies $T_a > T - \frac{1}{N}$.

For (i), a marginal increase in a decreases both T_a and T_e. Egoists are now only willing to sort with strictly fewer than T' periods remaining. T_a remains strictly larger than $T - \frac{1}{N}$, so altruists are still unwilling to sort with fewer than T periods remaining. The only equilibrium is therefore for all players to defect.

For (ii), the new rule cannot sustain complete cooperation if:

$$\delta_0 < \frac{1-\delta}{\delta}\frac{b-a}{a}$$

An egoist who deviates defects in fewer periods than an egoist who sorts.

If $\delta_0 \leq \lambda$ as well, an egoist who sorts and one who makes the same deviation as before both leave the game with probability one. That reduces expected pay-off in each period after by shock by a if an egoist sorts and by d if he deviates, and so increases the incentive to deviate.

If $\delta_0 > \lambda$, then the pay-off from deviating increases as the probability of leaving the game is now zero. The pay-off from sorting decreases because the probability of leaving the game is now higher: $\frac{1-\delta_0}{1-\lambda} > 1 - \delta_0$. Once again, deviating is more attractive.

Proof of Lemma 9.

Rearranging the result from Lemma 5 gives that in the general game, an altruist is willing to sort T periods before the shock if and only if $\lambda \geq \mu_a^T$, where:

$$\mu_a^T \equiv \frac{d - c - \frac{\delta}{1-\delta}(1 - \delta^{T-1})(a + \theta_0 - d)}{a + \theta_0 + d - b - c}$$

Given that cooperating will yield $a + \theta_0$ in all subsequent periods, he will only cooperate if there is at least a probability μ_a^T his partner will cooperate as well.

Let $0 \leq \mu_i < \mu_a^T$ and say there are T periods until the shock. I will show that both player i and j defect regardless of type in period T. The matching rule will then keep them together and the belief μ_i will remain unchanged. The condition $\mu_i < \mu_a^{T-1}$ will also hold, because μ_a^{T-1} is decreasing in T'. Player i and j will therefore defect regardless of type until the shock.

Say player i would defect if he were an egoist. Player j then believes the probability player i cooperates is less than or equal to μ_i. Player j must defect if he were an altruist because $\mu_i < \mu_a^T$. He must then also defect if he were an egoist, since cooperating would be a dominated action and would also reveal his type.

Player i is therefore sure that his partner will defect. Cooperation is only profitable for an altruist if there is at least a probability $\mu_a^T > 0$ his partner will cooperate, which is not the case. So player i will defect regardless of type.

Now say player i would cooperate if he were an egoist. Then he must must also cooperate if he were an altruist. The belief μ_i would remain unchanged in the next period, while the new critical value μ_a^{T-1} would now be larger. There must be some period where player i defects if he were an egoist. In that period, μ_i is lower than the critical value so both player i and his partner must defect regardless of type by the previous argument. That means player i would have an incentive to defect as an egoist in the previous period, and repeatedly applying the argument results in a contradiction.

Proof of Lemma 10.

An egoist's expected punishment is increasing in the probability he remains in the game in future periods. This probability, for some period $T_0 + M - 1$ with $M \leq N$, is:

$$(1-p)^M + p\delta_0 \sum_{i=0}^{M-1} (1-p)^i$$

The derivative of this expression with respect to p is:

$$-M(1-p)^{M-1} + \delta_0 \sum_{i=0}^{M-1} (1-p)^i - p\delta_0 \sum_{i=1}^{M-1} i(1-p)^{i-1}$$

I now show by induction that the derivative equals $-M(1-p)^{M-1}(1-\delta_0) < 0$. Taking $M = 1$ gives $-1(1-\delta_0)$. The derivative for $M+1$ is:

$$-(M+1)(1-p)^M + \delta_0 \sum_{i=0}^{M}(1-p)^i - p\delta_0 \sum_{i=1}^{M} i(1-p)^{i-1}$$

Using the induction hypothesis, this equals:

$$-M(1-p)^{M-1}(1-\delta_0)+M(1-p)^{M-1}-(M+1)(1-p)^M+\delta_0(1-p)^M-p\delta_0 M(1-p)^{M-1}$$

which simplifies to $-(M+1)(1-p)^M[1-\delta_0]$

Proof of Lemma 2

A sorting equilibrium does not exist for $M = 1$. But if $\tau < \frac{1}{M}$ and $\delta = 1$, then the incentive constraints for sorting are independent of M:

$$\frac{b}{N} + (1-\lambda)\frac{d}{N} - \lambda\frac{(a+\theta_0)}{N} - (1-\lambda)\frac{c}{N} \leq T\frac{(a+\theta_0-d)}{N}$$

$$\frac{b}{N} + (1-\lambda)\frac{d}{N} - \lambda\frac{a}{N} - (1-\lambda)\frac{c}{N} \leq T\frac{(a-d)}{N} + \frac{b-a}{N}$$

Now say $\tau > \frac{1}{M}$. Let $k(N)$ be the smallest integer such that $\frac{k(N)}{N} > \tau$. Actions are revealed $k(N)$ periods or time $\frac{k(N)}{N}$ after they occur. As N increases, $k(N)$ increases such that $\frac{k(N)}{N}$ tends to τ.

Choose N sufficiently large so that an integer M exists with $T_a < \frac{M}{k(N)} \leq T_e$. Fix N at such a value and refer to $k(N)$ as k. Note that $M \geq 2k$ since $T_e > 2$.

Say players pool until M periods before the shock, after which altruists always cooperate and egoists defect. Players update their beliefs after k periods and then rematch. An egoist's equilibrium pay-off in the last M periods is:

$$k\lambda\frac{b}{N} + k(1-\lambda)\frac{d}{N} + (M-k)\frac{d}{N}$$

I now show that it is sufficient to consider the following deviation: an egoist cooperates for k periods, continues cooperating with an altruist, then defects for the last k periods before the shock.

If an egoist defects as part of any deviation, then it is optimal to keep on defecting. Any cooperation will not prevent him from eventually being revealed as a past defector and punished.

An egoist may deviate by cooperating for the first $K < N$ periods where players sort. Others initially believe he is an altruist so he matches with an altruist partner after playing N periods. He then defects for K periods, at which point his initial defection is revealed. He matches up with an egoist and defects until the shock. The deviation gives pay-off:

$$\frac{K}{N}[\lambda a + (1-\lambda)c] + (1-\frac{K}{N})[\lambda b + (1-\lambda)d] + \frac{K}{N}b + (\frac{M}{N} - 1 - \frac{K}{N})d$$

This expression is linear in K, so it is optimal to choose $K = 0$ or $K = N-1$. The first case is just the egoist's equilibrium strategy, and in the second case it is even better to choose $K = N$ so he is not revealed as a deviant until the shock.

So the optimal deviation yields:

$$k\lambda\frac{a}{N} + k(1-\lambda)\frac{c}{N} + (M-2k)\frac{a}{N} + k\frac{b}{N}$$

Multiplying both expressions by $\frac{N}{k}$ and comparing them gives the incentive constraint of the original model but with T replaced by $\frac{M}{k}$. The egoist is willing to sort because $\frac{M}{k} \leq T_e$.

A parallel argument shows an altruist has no profitable deviation.

References

Alchian, A. A. and H. Demsetz (1972, December). Production , information costs, and economic organization. *American Economic Review 62*(5), 777–95.

Arrow, K. J. (1974). *The Limits of Organization*. New York, London: W. W. Norton and Company.

Barker, J. R. (1993). Tightening the iron cage: Concertive control in self-managing teams. *Administrative Science Quarterly 38*, 408–437.

Bernheim, B. D., B. Peleg, and M. D. Whinston (1987, June). Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory 42*(1), 1–12.

Bewley, T. F. (1999). *Why Wages Don't Fall during a Recession* (Third ed.). Harvard University Press.

Burt, R. S. (2005). *Brokerage and Closure: An Introduction to Social Capital*. Oxford, New York: Oxford University Press.

Che, Y.-K. and S.-W. Yoo (2001, June). Optimal incentives for teams. *American Economic Review 91*(3), 525–541.

Clegg, C., T. Wall, K. Pepper, C. Stride, D. Woods, D. Morrison, J. Cordery, P. Couchman, R. Badham, C. Kuenzler, G. Grote, W. Ide, M. Takahashi, and K. Kogi (2002). An international survey of the use and effectiveness of modern manufacturing practices. *Human Factors and Ergonomics in Manufacturing 12*(2), 171–191.

Faíña Medín, A., I. García-Jurado, J. Méndez-Naya, and L. Méndez-Naya (1998, December). Unilateral commitment in the finitely repeated prisoners' dilemma. *Annals of Operations Research 84*(0), 187–194.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. The MIT Press.

Gibbons, R. (1998, August). Incentives in organizations. *The Journal of Economic Perspectives 12*(4), 115–132.

Gneezy, U. and A. Rustichini (2000, August). Pay enough or don't pay at all. *The Quarterly Journal of Economics 115*(3), 791–810.

Hirshlifer, D. and E. Rassmusen (1989, August). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavior & Organization 12*(1), 87–106.

Holmstrom, B. and P. Milgrom (1990, March). Regulating trade among agents. *Journal of Organizational and Theoretical Economics 146*(1), 85–105.

James, D. and W. S. R. D. Costa (1994, December). Nash equilibrium under knightian uncertainty: Breaking down backward induction. *Journal of Economic Theory 64*(2), 305–324.

Janssen, M. C., J. Gorter, and S. van de Meerendonk (1997, April). Cooperation in a modified version of the finitely repeated prisoners' dilemma game. *Journal of Economic Behavior & Organization 32*(4), 613–619.

Janssen, M. C. and E. Mendys (2004, November). The price of a price: On the crowding out of social norms. *Journal of Economic Behavior & Organization 55*(3), 377–395.

Kandel, E. and E. P. Lazear (1992, August). Peer pressure and partnerships. *Journal of Political Economy 100*(4), 801–17.

Kinsley, B., C. Alpin, J. Forth, A. Bryson, H. Bewley, G. Dix, and S. Oxenbridge (2005). Inside the workplace: First findings from the 2004 workplace employment relations survey. *Department of Trade and Industry, Great Britain.*

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982, August). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory 27*(2), 245–252.

Kvaloy, O. and T. E. Olsen (2006, January). Team incentives in relational employment contracts. *Journal of Labor Economics 24*(1), 139–170.

Levin, J. (2002, August). Multilateral contracting and the employment relationship. *The Quarterly Journal of Economics 117*(3), 1075–1103.

Osterman, P. (2000, January). Work reorganization in an era of restructuring: Trends in diffusion and effects on employee welfare. *Industrial and Labor Relations Review 53*(2), 179–196.

Rayo, L. (2007). Relational incentives and moral hazard in teams. *Review of Economic Studies 74*(3), 937–963.

Rotemberg, J. J. (1994, August). Human relations in the workplace. *Journal of Political Economy 102*(4), 684–717.

Ryan, A. M., M. J. Schmit, and R. Johnson (1996). Attitudes and effectiveness: Examining relations at an organizational level. *Personnel Psychology 49*(4), 853–882.

Serrano, R. and I. Zapater (1998, February). The three-legged race: Cooperating to compete. *Games and Economic Behavior 22*(2), 343–363.

Uzzi, B. (1996, August). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review 61*(4), 674–698.

Wood, S., C. Stride, T. Wall, and C. Clegg (2004). Revisiting the use and effectiveness of modern management practices. *Human Factors and Ergonomics in Manufacturing 14*(4), 415–432.