



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Sorting and Sustaining Cooperation

Citation for published version:

Vikander, N 2012, 'Sorting and Sustaining Cooperation', *Oxford Economic Papers*, vol. 65, no. 2, pp. 548-566. <https://doi.org/10.1093/oep/gps030>

Digital Object Identifier (DOI):

[10.1093/oep/gps030](https://doi.org/10.1093/oep/gps030)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Oxford Economic Papers

Publisher Rights Statement:

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in Oxford Economic Papers following peer review. The definitive publisher-authenticated version © Vikander, N. (2012). Sorting and Sustaining Cooperation. Oxford Economic Papers, doi: 10.1093/oep/gps030 is available online at: <http://oep.oxfordjournals.org/content/65/2/548.abstract>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Sorting and sustaining cooperation

Nick Vikander

School of Economics, University of Edinburgh. 21 Buccleuch Place, EH8 9LN;
email: nick.vikander@ed.ac.uk

Abstract

This paper looks at cooperation in teams where some people are selfish and others are conditional cooperators, and where lay-offs will occur at a fixed future date. I show that the best way to sustain cooperation prior to the lay-offs is often in a sorting equilibrium, where conditional cooperators can identify and then work with one another. Changes to parameters that would seem to make cooperation more attractive, such as an increase in the discount factor or the fraction of conditional cooperators, can reduce equilibrium cooperation if they decrease a selfish player's incentive to sort.

JEL classifications: L23, D82, M50

Keywords: team production, lay-offs, sorting, conditional cooperator

1 Introduction

Managers are often preoccupied with sustaining high morale in the workplace. Bewley (1999) reports their widely-held view that employee morale is important for productivity, but that high morale also tends to be fragile. In particular, increased turnover or expected lay-offs can cause morale to break down.

To gain insight into the question of work morale, Gächter (2006) argues a key point is to recognize that people differ in their intrinsic motivation to cooperate. Some people are selfish and are willing to free-ride on the work of others. Yet a wealth of evidence suggests many other people are conditional cooperators, who will cooperate if they expect others to do the same (see e.g. Keser and van Winden (2000), Fischbacher et al. (2001), Frey and Meier (2004)). Conditional cooperation is widespread, with many experimental estimates of the proportion of conditional cooperators ranging from 40 to 60 % (see Chaudhuri (2011)).

The presence of conditional cooperators can help explain cooperation when free-riding is possible, while the presence of selfish types in a heterogeneous population can shed light on how cooperation may break

down over time. In this sense, Fehr and Schmidt (2001) argue that ‘[t]he interaction between fair and selfish people is key to understanding ... observed behavior in strategic settings.’

This paper takes up the issue of cooperation in teams in the face of upcoming lay-offs, where some players are egoists, others are conditional cooperators and where type is private information. There is an infinite time horizon and players pair up into teams to play a stage game in each period. All players know that lay-offs will occur at the end of period T_0 , at which point a given fraction of players will leave the game.

The stage game has players choose between a selfish and a cooperative action. Egoists have a dominant strategy to take the selfish action, whereas conditional cooperators prefer to cooperate if they expect their teammates to do the same. A rematching mechanism allows conditional cooperators to rematch together if they reveal their type through earlier play. The incentives to cooperate arise both through intrinsic motivation and through repeated interactions, and I explore to what extent cooperation is possible leading up to the lay-offs.

I show that when the fraction of conditional cooperators is small and the size of the lay-offs is large, then the best way to sustain cooperation will be through sorting. In a sorting equilibrium, players reveal their type by taking different actions in some period $t \leq T_0 - 1$, and rematching then allows conditional cooperators to work together up until the lay-offs. I derive necessary and sufficient conditions for the existence of a sorting equilibrium, and show that one will exist whenever the material pay-off to unreciprocated cooperation is low and the intrinsic motivation of conditional cooperators is high.

I also show that an increase in parameters that would seem to favour cooperation, such as the discount factor or the fraction of conditional cooperators, can decrease the amount of cooperation in equilibrium. The relationship between these parameters and equilibrium cooperation will sometimes be non-monotonic. Further results suggest that a firm may benefit from increasing the notice it gives about upcoming lay-offs, and that laying off workers who reveal they are ‘bad apples’ can be counterproductive.

Sorting is not helpful when lay-offs are small, as a pooling equilibrium then exists with full cooperation in all periods. Given some assumptions on the discount factor, egoists will cooperate because they are likely to remain in the game after period T_0 to be punished following a deviation. If lay-offs are large but the fraction of conditional cooperators is large as well, then almost full cooperation is still possible in a non-sorting equilibrium. Conditional cooperators will cooperate in all periods, while egoists will only defect in period T_0 . Egoists wait to defect until just before the lay-offs, because they know their cooperative teammates would withdraw all cooperation following an earlier deviation.

Cooperation will otherwise break down in any non-sorting equilibrium. Egoists eventually defect because they are likely to leave the game after period T_0 , and conditional cooperators will defect to prevent egoists

from taking advantage of them. Players will then also defect in all earlier periods by a logic of backwards induction. In this case, more cooperation can be sustained in a sorting equilibrium where conditional cooperators reveal their type in some period $t \leq T_0 - 1$, rematch and cooperate for $T_0 - t$ periods up until the lay-offs.

For a sorting equilibrium to exist, two incentive constraints must hold. There must be some period where conditional cooperators are willing to cooperate even though egoists defect, and where egoists will defect even though conditional cooperators cooperate. These constraints will hold when intrinsic motivation is high and the material pay-off to unreciprocated cooperation is low, since this drives a wedge between the expected pay-off from cooperating for the two different types.

A parameter change that would seem to favour cooperation can have the opposite effect if it causes a sorting equilibrium to break down. When players are more patient, or conditional cooperators more numerous, then an egoist may prefer to deviate from a sorting equilibrium by imitating a conditional cooperator and later taking advantage of him. Promising to fire players who are believed to be selfish (bad apples) can also reduce equilibrium cooperation by decreasing an egoist's incentive to sort.

The experimental literature on public goods games has shown that cooperation depends crucially on group composition. As in this paper, cooperation tends to increase when conditional cooperators become confident they are working with one another. Gächter and Thoni (2005) show that grouping conditional cooperators together will increase subsequent cooperation, and Burlando and Guala (2005) demonstrate that similar results are robust to different ways of classifying types. Gunthorsdottir et al. (2007) show that rematching players based on previous play can help, even if players do not know the rematching rule. de Oliveira et al. (2009) also indicate that conditional cooperators who are matched together will cooperate more still if they are informed about this fact. A major difference is that I look at cooperation in the shadow of upcoming lay-offs and where the timing of sorting is endogenous, whereas the timing of changes to group composition in most experimental work is exogenous.

Teams are a relevant setting to look at issues of cooperation, rematching, and group composition because of the rise of self-managed work teams. Using data from Lawler et al. (1995) and Lawler et al. (2001), Lazear and Shaw (2007) show that the percentage of large firms using self-managed work teams rose from 27 to 78 % from 1987-96. These self-managed teams themselves decide on group composition and whether to exclude team members who shirk (Lawler and Cohen, 1992). Barker (1993) illustrates this process with an ethnographic study of a small manufacturing company, where workers decide on team membership.

This paper also relates to recent work on incentives within the firm when workers have social preferences. Bartling and von Siemens (2010a) show that an equal sharing rule in partnerships can be optimal if workers

are envious of one another, and Bartling and von Siemens (2010b) examine how worker envy can affect incentives under moral hazard. Some experimental work also suggests that a firm might sort its own employees based on social preferences, as selfish types tend to prefer tournament-based variable pay (Dohmen and Falk, 2011).

Other papers have looked at how workers with different social preferences may sort between different firms. Kosfeld and von Siemens (2009) and Kosfeld and von Siemens (2011) show that when some workers are conditional cooperators and others are selfish, then a separating equilibrium may exist where cooperative types all apply to the same firms. These firms can make positive profits despite free entry since a higher wage would only attract selfish workers. Firms may also want to screen workers based on other social preferences such as inequity aversion (von Siemens, 2011) or intrinsic motivation to work (Delfgaauw and Dur 2007, 2008, 2009). One difference here is that I do not consider the sorting of workers between different firms, but between different teams. Another is that interactions are repeated, so that future punishments can influence cooperation.

In this paper, the strategic situation prior to the lay-offs resembles a finitely repeated prisoners' dilemma. As in Kreps et al. (1982) and Conlon (2003), some players are behavioral types and reputation is important. Some players will remain in the game after the lay-offs, however, which affects incentives for earlier cooperation. More broadly, this paper differs in its focus. Kreps et al. (1982) examine whether the standard theoretical predictions for the finitely repeated prisoners' dilemma are robust to the introduction of a small number of behavioral types. This paper instead follows in the spirit of the earlier quote from Fehr and Schmidt (2001). It looks at the interaction between selfish and intrinsically motivated types, both of which may be numerous, to understand how cooperation evolves in the shadow of upcoming lay-offs.

The rest of the paper is organized as follows. Section 2 lays out the model, and Section 3 contains the main analysis on how sorting can help sustain cooperation. Section 4 looks at comparative statics and equilibrium cooperation. Section 5 examines the impact of linking lay-offs to performance, and Section 6 examines issues of robustness. Section 7 then concludes. All proofs can be found in the appendix.

2 The Model

There are a countably infinite number of players, indexed by $i \in \mathbb{N}$. Players differ in type $\theta \in \{\theta_E, \theta_C\}$, where $\theta_i = \theta_E$ if player i is an egoist and $\theta_i = \theta_C$ if he is a conditional cooperator. The ex-ante probability that any player is a conditional cooperator is $\lambda \in (0, 1)$, and type is private information. Time is discrete with an infinite horizon and indexed by $t = 1, 2, \dots$. Players have a common discount factor $\delta \in (0, 1]$.

In each period, players work in teams of two on a productive task. Teammates play a stage game, where each player can choose either to cooperate (C) or to defect (D). If i is the row player and j the column player, pay-offs are

	C	D
C	$a + \Gamma_i, a + \Gamma_j$	c, b
D	b, c	d, d

The terms a, b, c and d are the material pay-offs, where $b > a > d > c > 0$, and $2a > b + c$. Following the notation of Kosfeld and von Siemens (2011), player i 's extra intrinsic utility from reciprocated cooperation is $\Gamma_i = 0$ if $\theta_i = \theta_E$ and $\Gamma_i = \Gamma > b - a$ if $\theta_i = \theta_C$. Egoists only value material incentives, and so they face a prisoners' dilemma. Conditional cooperators enjoy intrinsic utility from reciprocated cooperation, so conditional cooperators who knew each others' type would face a coordination game. I denote player i 's period t action by $a_{it} \in \{C, D\}$, and the vector of all period t actions by $\mathbf{a}_t = \{a_{it}\}_{i \in \mathbb{N}}$.

Players know that lay-offs will occur at the end of period T_0 , at which time a randomly selected group of players will leave the game. The probability that any given player will be laid off is $1 - \delta_0$, where $0 \leq \delta_0 < 1$. A player who is laid off has a pay-off of zero in all subsequent periods.

If players i and j are teammates then I say that they are matched together. Let m_{it} denote the player with whom i is matched at the start of period t , where $m_{it} = j$ means that players i and j play the stage game in period t . Let the vector $\mathbf{m}_t = \{m_{it}\}_{i \in \mathbb{N}}$ describe the set of matches, or teams, at the start of period t .

When choosing his period t action a_{it} , player i observes the full history of play h_{t-1} up to and including period $t - 1$, where $h_0 = \mathbf{m}_1$ and $h_t = h_{t-1} \times \mathbf{a}_t \times \mathbf{m}_{t+1}$. A pure strategy s_i for player i is a function which, for any history h_t , selects an action $a_{it} \in \{C, D\}$.

Let $\mu_i(j|h_{t-1}) \in [0, 1]$ denote the belief of player i at the start of period t that any other player j is a conditional cooperator. Since histories are fully observable, players hold the same beliefs and I can drop the subscript i . I simply write μ_{jt} , where the dependence on a particular history is implicit.

Players are randomly matched into teams before the start of period 1. Each team continues to work together in all subsequent periods, unless one player is laid off or unless there is some period t' where one player cooperates while his teammate defects. The interpretation in the latter case is that a team splits up when one teammate has taken advantage of the other.

Any player whose team splits up at the end of period t' is randomly rematched with another player whose team has also split up in that period, and who took the same period t' action: $m_{it'+1} = j$ if and only if $a_{it'} = a_{jt'}$. If this rematching is infeasible, say because the number of teams that split up is odd, then one

player is randomly matched to another who took a different period t' action, and the remaining players are rematched as above.

This exogenous matching mechanism reflects the idea that players prefer to work with teammates who cooperate and that working together requires mutual consent. The results will not depend on the fine details of the matching mechanism, such as whether a team should also split up after both players defect. What is relevant is that on the equilibrium path, players who cooperate in period t will be able to work together in period $t+1$. I also argue in Section 6 that qualitatively similar results will continue to hold if players cannot rematch.

Rematching means that observability is important, since each player should know the past actions of his current teammate. Full observability may be more plausible when there are relatively few players. Having a small number of players would make the exposition less clear, but all the qualitative results would continue to hold.

I look for symmetric, pure strategy Perfect Bayesian equilibria, where all players of the same type use the same strategy. I denote the strategy of conditional cooperators by s_C and the strategy of egoists by s_E . Each player's strategy must be optimal given the strategies of other players and given beliefs. Beliefs are consistent with player strategies, in the sense of following from Bayes' rule whenever possible.

I am concerned with how upcoming lay-offs can cause cooperation to break down. To separate this effect from simple impatience, I assume that in the absence of lay-offs players could sustain cooperation in all periods

$$\delta \geq \frac{b-a}{b-d}. \quad (1)$$

When there are multiple equilibria, I will be interested in the equilibrium that involves the most cooperation. By this I mean the equilibrium that maximizes $\sum_{t=1}^{\infty} x_t$, where x_t is the fraction of players who cooperate in period t .

3 Analysis

Throughout the analysis, I will assume that players use trigger strategies with the following feature. Consider a particular candidate equilibrium, and suppose that some player i takes an action $a_{it'}$ after history $h_{t'}$ that neither s_C nor s_E prescribe for player i after this particular history. Suppose furthermore that t' is the first period where any player has taken such an unexpected action. Then both s_C and s_E call on a player to defect in any later period $t \geq t' + 1$ where he is matched with player i .

With these trigger strategies, a player who first deviates from a candidate equilibrium by taking an unexpected action will be punished in all later periods. The punishment is self-enforcing because mutual defection is a Nash equilibrium of the stage game for both types. It also satisfies the requirements of Perfect Bayesian equilibrium for any out-of-equilibrium beliefs, and in that sense it is robust. Trigger strategies allow me to derive the maximum amount of cooperation that can be sustained in equilibrium, and then focus on the impact of sorting.

The results will not depend on any further details about play off the equilibrium path, other than what is specified above. These trigger strategies specify how a player will be punished if he is the first to deviate from a candidate equilibrium by taking an unexpected action, and the issue is whether any player wants to be the first to deviate. For the sake of convenience, I therefore assume the simplest self-enforcing punishment strategy. Following an unexpected action by any player in period t' , s_C and s_E have all players defect in all later periods $t \geq t' + 1$.

This punishment simplifies the exposition as I can then describe any equilibrium strategy just by its actions on the equilibrium path. It also avoids the question of exactly what other play may also be possible off the equilibrium path, after a particular sequence of unexpected actions and for a particular set of out-of-equilibrium beliefs. The reasoning is not that cooperation must necessarily break down in all teams following an unexpected action by a single player. Rather, we can remain agnostic about the extent of cooperation off the equilibrium path for all players other than the first to deviate, as it will have no bearing on the results.

The first result shows that it is impossible to sustain full cooperation in all periods unless δ_0 exceeds a certain threshold, so unless the fraction of players laid off is sufficiently small.

Proposition 1. *An equilibrium exists where both types cooperate in all periods if and only if $\delta_0 \in [\delta_0^*, 1]$, where*

$$\delta_0^* = \frac{1 - \delta}{\delta} \frac{b - a}{a - d}.$$

The intuition for the result is straightforward. If there were no lay-offs, $\delta_0 = 1$, then (1) implies players would be patient enough to cooperate in all periods. If instead lay-offs are sufficiently large, $\delta_0 < \delta_0^*$, then an egoist who deviates by defecting before period T_0 knows he is unlikely to remain in the game to be punished. His most attractive deviation by (1) is to wait until period T_0 before defecting, and if he is laid off escape the punishment altogether. The critical value δ_0^* is decreasing in the discount factor because a player who is patient is less tempted to deviate.

Although full cooperation is impossible when $\delta_0 < \delta_0^*$, cooperation after the lay-offs can still be sustained. In any candidate equilibrium, (1) implies each player is patient enough to cooperate for all $t \geq T_0 + 1$ if he

expects other players to cooperate as well. From now on, I will only consider candidate equilibria where, on the equilibrium path, both types cooperate in all periods after the lay-offs. I can then examine how much cooperation is also possible before period T_0 . Doing so focuses the analysis on how cooperation may break down due to upcoming lay-offs, not due to simple miscoordination on an inefficient equilibrium.

For brevity, I will often describe a candidate equilibrium simply in terms of actions prior to period T_0 . The implicit understanding should then be that both types cooperate on the equilibrium path for all $t \geq T_0 + 1$.

To explore how cooperation can be sustained prior to the lay-offs when $\delta_0 < \delta_0^*$, I introduce the notion of a sorting equilibrium.

Definition. An equilibrium is a *sorting equilibrium* if, on the equilibrium path, conditional cooperators and egoists take different actions in at least one period $t \leq T_0 - 1$.

In a sorting equilibrium, different types take different actions at some point before the last period prior to the lay-offs. I describe the first such period t as the period where players sort, and refer to any other equilibrium as a non-sorting equilibrium.

This definition implies that an equilibrium where different types first take different actions in period T_0 itself is a non-sorting equilibrium. With sorting, I want to focus on how conditional cooperators can identify each other, rematch and then cooperate in all later periods until the lay-offs. This later cooperation cannot occur if players first choose different actions in period T_0 , since the lay-offs then follow immediately.

The following result describes the maximum amount of cooperation that can be sustained in any non-sorting equilibrium. It also establishes conditions under which sorting will unambiguously help, so where any sorting equilibrium will involve more cooperation than every non-sorting equilibrium.

Proposition 2. *Suppose $\delta_0 < \delta_0^*$, as given in Proposition 1. If both of the following conditions hold*

$$\lambda \geq \frac{d - c}{a + \Gamma - b - c + d}, \quad (2)$$

$$\lambda \geq \frac{1}{\delta(b - d)} \left((b - a) - \delta_0 \frac{\delta^2}{1 - \delta} (a - d) \right), \quad (3)$$

then the equilibrium with the most cooperation will be a non-sorting equilibrium. Otherwise, any sorting equilibrium will involve more cooperation than every non-sorting equilibrium.

If (2) and (3) hold, then the non-sorting equilibrium with the most cooperation has both types cooperate for all $t \leq T_0$ except for egoists who defect in period T_0 . If (3) is violated but (2) is not, then it instead has both types defect for all $t \leq T_0$ except for conditional cooperators who cooperate in period T_0 . If (2) is violated, then it has both types defect for all $t \leq T_0$.

Sorting is not necessary to sustain cooperation, even with large lay-offs, if the fraction of conditional cooperators is also large. When (2) and (3) hold, a non-sorting equilibrium will exist where both types cooperate in all periods, except for egoists who defect immediately before the lay-offs.

The result suggests that a firm with enough intrinsically motivated workers can always sustain cooperation in the face of lay-offs, but I will show in Section 4 that the relationship between λ and equilibrium cooperation need not be monotonic. If teammates' decisions to cooperate are strategic complements, then a small increase in the fraction of conditional cooperators can actually reduce equilibrium cooperation. Proposition 2 also implies that whenever the fraction of conditional cooperators is below a certain threshold, sorting will be crucial for sustaining cooperation.

Condition (2) refers to a conditional cooperator's incentive to cooperate in period T_0 , where he expects other conditional cooperators to cooperate and egoists to defect. A conditional cooperator who deviates will not be punished because defecting allows him to mimic an egoist. He will cooperate if cooperating gives a higher immediate pay-off than defecting, which is the case if his teammate is likely a conditional cooperator as well.

If (2) is violated, then all cooperation prior to the lay-offs will break down in any non-sorting equilibrium. Egoists will defect in period T_0 by $\delta_0 < \delta_0^*$, as will conditional cooperators since (2) does not hold. Egoists foresee that nobody will cooperate in period T_0 so they will also defect in period $T_0 - 1$, and any non-sorting equilibrium must have conditional cooperators choose the same period $T_0 - 1$ action. Following this logic of backwards induction means that no cooperation is possible until after the lay-offs.

Condition (2) may be necessary to sustain any appreciable amount of cooperation in a non-sorting equilibrium, but it is not sufficient. Condition (3) refers to an egoist's incentive to cooperate in period $T_0 - 1$, when he expects all players to cooperate and then egoists to defect in period T_0 . An egoist who deviates by defecting in period $T_0 - 1$ will be punished in period T_0 , but this punishment can only reduce his pay-off if his teammate turns out to be a conditional cooperator. The worst his period T_0 teammate can do is to defect, which an egoist would also have done in the candidate equilibrium. A large fraction of conditional cooperators makes the expected punishment more severe, which discourages an egoist from defecting earlier than expected and helps to sustain cooperation for all $t \leq T_0 - 1$.

Any sorting equilibrium will involve more cooperation than every non-sorting equilibrium when either (2) or (3) is violated, because sorting allows conditional cooperators to cooperate in multiple periods before the lay-offs. Sorting can only occur in a period where conditional cooperators cooperate and egoists defect. If instead egoists cooperated, then an egoist who deviated could increase his immediate pay-off and benefit from rematching with a conditional cooperator. After players sort in period t' , the matching rule pairs

conditional cooperators together in period $t' + 1$. They can then continue cooperating for all $T_0 - t'$ periods until the lay-offs, where $T_0 - t' \geq 1$.

The following result describes necessary and sufficient conditions for the existence of a sorting equilibrium.

Proposition 3. *A sorting equilibrium where different types first take different actions in period $t' \leq T_0 - 1$ will exist if and only if*

$$\lambda b + (1 - \lambda)d - \lambda(a + \Gamma) - (1 - \lambda)c \leq \frac{\delta}{1 - \delta}(1 - \delta^{T_0 - t'})(a + \Gamma - d), \quad (4)$$

$$\frac{\delta}{1 - \delta}(1 - \delta^{T_0 - t' - 1})(a - d) + \delta^{T_0 - t'}(b - d) - \delta_0 \frac{\delta^{T_0 - t' + 1}}{1 - \delta}(a - d) \leq \lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c. \quad (5)$$

A sufficient condition for (4) and (5) is that the intrinsic motivation of conditional cooperators is sufficiently high and the pay-off to unreciprocated cooperation is sufficiently low: $\Gamma \geq K_1$ and $d - c \geq K_2$, where the critical value K_1 may depend on c , but K_2 does not depend on Γ .

When deciding whether to cooperate in period t' , a conditional cooperator must weigh the future benefit of following his equilibrium strategy against its immediate cost. The future benefit comes from revealing his type. Cooperating in period t' allows him to rematch with another conditional cooperator in period $t' + 1$ and earn $a + \Gamma$ by cooperating until the lay-offs. The immediate cost of cooperating in period t' is that his teammate may be an egoist who will defect, giving the lowest pay-off of c . A conditional cooperator can avoid this pay-off by defecting in period t' , but this deviation will cost him future cooperation as players believe he is an egoist. He must then rematch with an egoist in period $t' + 1$ and earn only $d < a + \Gamma$ by defecting until the lay-offs.

The right-hand side of (4) gives the future benefit to a conditional cooperator from playing his equilibrium strategy in period t' and the left-hand side gives the immediate cost. The future benefit is decreasing in t' , which shows that conditional cooperators will only sort if the lay-offs are sufficiently far off.

An egoist has different incentives in period t' , as defecting involves a future cost and an immediate benefit. The future cost comes from revealing that he is an egoist, which makes further cooperation before the lay-offs impossible. He is then rematched with another egoist in period $t' + 1$ and earns d in each period. In contrast, defecting in period t' provides an immediate benefit because it is an egoist's dominant action in the stage game. An egoist can sacrifice some immediate pay-off by cooperating in period t' , but this deviation means he rematches with a conditional cooperator and can earn $a > d$ in each future period. He can in fact do better still by cooperating until period T_0 and defecting just before the lay-offs, even though he will be punished for this unexpected action if he remains in the game.

The left-hand side of (5) gives the future cost to an egoist of playing his equilibrium strategy in period t' and the right-hand side gives the immediate benefit. The future cost is decreasing in t' , so that egoists will only sort if the upcoming lay-offs are sufficiently close by.

A sorting equilibrium will exist whenever the intrinsic motivation of conditional cooperators is sufficiently high and the pay-off to unreciprocated cooperation is sufficiently low. If the difference between c and d is large, then an egoist will defect in period t' because cooperation simply provides too low an expected material pay-off. A conditional cooperator will still cooperate as long as he obtains a sufficiently high intrinsic pay-off from reciprocated cooperation. A large value of Γ increases the wedge between the returns to cooperation for different types and can allow players to sort.

The timing of sorting is a balancing act, as lay-offs must be sufficiently close by to satisfy egoists and sufficiently far off to satisfy conditional cooperators. One implication of Proposition 3 is that a firm may sometimes be able to benefit from announcing lay-offs farther in advance. This advance notice can help sustain cooperation if sorting then gives conditional cooperators enough future benefit to reveal their type.

Having established when a sorting equilibrium will exist, the following result shows precisely how much cooperation is possible in a sorting equilibrium.

Proposition 4. *Suppose that a sorting equilibrium exists, and let t'_H and t'_L be the largest and smallest values of t' that satisfy both (4) and (5). Consider*

$$\lambda \geq \frac{1}{\delta(b-d)} \left((b-a) - \delta_0 \frac{\delta^{T_0-t'+2}}{1-\delta} (a-d) \right). \quad (6)$$

If t'_H satisfies (6), then the sorting equilibrium with the most cooperation has conditional cooperators cooperate for all $t \leq T_0$, while egoists cooperate for all $t \leq t'_H - 1$ and defect for all $t'_H \leq t \leq T_0$.

If t'_H violates (6), then it instead has egoists defect for all $t \leq T_0$, while conditional cooperators defect for all $t \leq t'_L - 1$ and cooperate for all $t'_L \leq t \leq T_0$.

It is straightforward to describe the amount of cooperation that is possible after players sort. Conditional cooperators can cooperate in all periods because they are matched together, while egoists must defect until the lay-offs because they have revealed their type. Proposition 4 shows that the fraction of conditional cooperators will determine whether cooperation is also possible before sorting.

If $t' = t'_H$ satisfies (6), then players can cooperate in all periods before they sort. The sorting equilibrium with the most cooperation has players sort late in the game, $t' = t'_H$, so that egoists cooperate for as long as possible before revealing their type. Both types cooperate until period t'_H and then egoists defect until the lay-offs.

Cooperation is otherwise impossible before players sort, so the sorting equilibrium with the most cooperation has players sort early in the game, $t' = t'_L$. Conditional cooperators can then reveal their type and begin to cooperate as soon as possible. Both types defect until period t'_L , at which point conditional cooperators begin to cooperate.

Cooperation before sorting depends on λ for a similar reason to condition (3) from Proposition 2. If an egoist expects all players to cooperate in period $t' - 1$ and then players to sort in period t' , he may be tempted to deviate by defecting a period earlier than expected. His teammate will then punish him by defecting in period t' , which is exactly what an egoist would also have done in the candidate equilibrium. The punishment only reduces a player's period t' pay-off if his teammate turns out to be a conditional cooperator, so the incentive to deviate is decreasing in λ .

Condition (6) is evaluated at $t' = t'_H$, rather than some lower value of t' , because sorting late in the game increases the weight of the punishment following an egoist's deviation. A player who deviates by defecting in period $t' - 1$ will be punished in all later periods. The punishment cannot reduce an egoist's pay-off for $t' + 1 \leq t \leq T_0$, since his teammate would also defect in the candidate equilibrium. In contrast, the punishment will reduce an egoist's pay-off from a to d in all periods after the lay-offs. The deviation is least attractive when the periods after the lay-offs are not heavily discounted from the perspective of period $t' - 1$, so when t' is close to T_0 . If any sorting equilibrium exists such that (6) is satisfied, then (6) will also be satisfied for $t' = t'_H$.

4 Comparative Statics and Equilibrium Cooperation

The importance of sorting means that changes in certain parameters can have unexpected effects on equilibrium cooperation. A change that would seem to make cooperation more attractive, such as an increase in the discount factor or the fraction of conditional cooperators, can reduce equilibrium cooperation if it leaves players unable to sort.

Making cooperation more attractive can cause sorting to break down because it tightens an egoist's incentive constraint, (5). An egoist's optimal deviation from a sorting equilibrium is to imitate a conditional cooperator, and then take advantage of his teammate after cooperating until period T_0 . Increasing for example the pay-off from reciprocated cooperation will make this deviation more attractive. If (5) becomes violated then a sorting equilibrium will break down, which by Proposition 2 can reduce equilibrium cooperation.

Drawing general conclusions about how parameter changes affect equilibrium cooperation is difficult,

because conditional cooperators and egoists have very different incentives in a sorting equilibrium. A change that makes period t' cooperation more attractive will loosen the incentive constraint of a conditional cooperator, (4), but tighten the incentive constraint of an egoist, (5). Both incentive constraints must hold in a sorting equilibrium, so the impact of a parameter change on equilibrium cooperation is often ambiguous.

For this reason, I turn to a related but more specific question: are there situations where the relationship between these parameters and the amount of equilibrium cooperation is non-monotonic? The reasoning is that a sufficiently large increase in the discount factor or the fraction of conditional cooperators could promote full cooperation in all periods, but a smaller increase might just prevent sorting. Propositions 4 and 5 establish conditions under which this is indeed the case.

I first state a lemma that will be useful in proving the results.

Lemma 1. *Consider (4) and (5) from Proposition 3. If (4) holds for t' , then it will also hold for all $t \leq t' - 1$. If (5) holds for t' , then it will also hold for all $t \geq t' + 1$.*

Any t' that satisfies (4) with equality will strictly satisfy (5), and any t' that satisfies (5) with equality will strictly satisfy (4).

The following proposition shows there are parameters for which the relationship between the discount factor and the amount of equilibrium cooperation is non-monotonic.

Proposition 5. *Suppose that $(b - a) - (d - c)$ is small but strictly positive, $b - a < (a - d)(1 - \delta_0)$ and $\delta_0 > 0$. Then for $a + \Gamma$ sufficiently close to b , the relationship between δ and the amount of equilibrium cooperation is non-monotonic. Specifically, there exist $\delta_1 < \delta_2 < \delta_3 < 1$ such that:*

For $\delta \in [\delta_1, \delta_2]$, the equilibrium with the most cooperation has players sort in period $t' = T_0 - 1$.

For $\delta \in (\delta_2, \delta_3)$, it has both types defect for all $t \leq T_0$.

For $\delta \in [\delta_3, 1]$, it has both types cooperate for all $t \leq T_0$.

When the discount factor is sufficiently high, $\delta \in [\delta_3, 1]$, Proposition 1 shows that an equilibrium will exist where players cooperate in all periods. If the discount factor drops to $\delta < \delta_3$, an egoist's deviation from this candidate equilibrium becomes profitable, and the parameter assumptions imply that (5) is also violated on an interval (δ_2, δ_3) . Egoists are unwilling to sort for these values of δ which rules out any cooperation before the lay-offs. When the discount factor is lower still, $\delta \in [\delta_1, \delta_2]$, then (5) becomes satisfied and sorting increases cooperation.

This non-monotonicity result holds under quite specific parameter assumptions. It may also be useful to say something more, to identify when an increase in the discount factor will clearly increase cooperation and when its impact will be ambiguous.

An increase in δ always reduces an egoist's incentive to deviate from a candidate equilibrium with full cooperation in all periods. Deviating from such a candidate equilibrium provides an egoist with an immediate benefit, but also carries a future cost as he is punished in later periods. An increase in the discount factor increases the weight of this cost.

An increase in δ also increases an egoist's incentive to deviate from any sorting equilibrium. Deviating from a sorting equilibrium provides a future benefit, since an egoist can mimic a conditional cooperator and later take advantage of him. An increase in the discount factor increases the weight of this benefit, tightening (5) and making the deviation more attractive.

The non-monotonicity result follows from these two opposing effects. These effects also imply that a sufficiently large increase in the discount factor to make $\delta_0^* \leq \delta_0$ will clearly help by making full cooperation possible. A smaller increase that leaves $\delta_0 < \delta_0^*$ will just increase an egoist's incentive to deviate from any sorting equilibrium, and its impact on cooperation will be ambiguous.

I now consider a similar question but with the fraction of conditional cooperators.

Proposition 6. *Suppose $\delta = 1$ and $\delta_0 = 0$. Then for $b - d < d - c$, and $a + \Gamma$ sufficiently close to b , the relationship between λ and the amount of equilibrium cooperation is non-monotonic. Specifically, there exist $\lambda_1 < \lambda_2 < \lambda_3 < 1$ such that:*

For $\lambda \in [\lambda_1, \lambda_2]$, the equilibrium with the most cooperation has players sort in period $t' = T_0 - 1$.

For $\lambda \in (\lambda_2, \lambda_3)$, it has both types defect for all $t \leq T_0$.

For $\lambda \in [\lambda_3, 1]$, it is a non-sorting equilibrium where conditional cooperators cooperate in at least period T_0 . Specifically, for λ sufficiently close to one, it has both types cooperate for all $t \leq T_0$ except for egoists who defect in period T_0 .

Proposition 2 showed that almost full cooperation can be sustained without sorting when the fraction of conditional cooperators is sufficiently high. If instead λ takes on a lower value, $\lambda \in (\lambda_2, \lambda_3)$, then these parameter values imply that cooperation breaks down because a sorting equilibrium does not exist. Cooperation again becomes possible when $\lambda \in [\lambda_1, \lambda_2]$, as (5) is then satisfied and egoists are willing to sort.

An important condition here is $b - d < a - c$, implied by $b - d < d - c$, which says that teammates' decisions to cooperate are strategic complements. Strategic complementarity means that an increase in λ will tighten the egoist incentive constraint (5) in any sorting equilibrium. An egoist deviates from a sorting equilibrium by cooperating, which under strategic complementarity is more attractive if his teammate will cooperate as well. An egoist will then have a larger incentive to deviate if his teammate is a conditional cooperator, which can cause a sorting equilibrium to break down.

In this way, the relationship between the fraction of conditional cooperators and equilibrium cooperation depends on the nature of production complementarities within the team. If teamwork is necessary for production but there are decreasing returns to effort within the team (strategic substitutes), then a firm can clearly benefit from increasing its proportion of intrinsically motivated workers. If instead there are increasing returns to effort within the team (strategic complements), then the impact can be ambiguous.

5 Linking Lay-offs to Performance

I have emphasized how sorting can help sustain cooperation in the face of upcoming lay-offs, but one limitation has been the assumption of a random firing rule. The probability of being laid off after period T_0 has been constant at $1 - \delta_0$, regardless of a player's actions in previous periods.

A random firing rule might be plausible if actions are difficult for a manager to observe, but it would otherwise seem natural to base lay-offs on performance. A manager might prefer to lay off workers who have produced low output or who have exhibited bad character. Basing lay-offs on performance would also be consistent with Bewley (1999), who argues that managers often want to fire workers who are bad apples.

I now examine the effect of a firing rule where bad apples are more likely to be laid off. I still assume that a fraction $1 - \delta_0$ of players are laid off after period T_0 , but priority is now given to those believed most likely to be egoists.

This performance-based firing rule will increase the range of parameters for which full cooperation in all periods is possible, as long as deviating increases the probability of being laid off. For this to be the case, the out-of-equilibrium belief about a deviating player being a conditional cooperator must be more negative than the prior. Such a belief is certainly reasonable as only an egoist could possibly benefit from a deviation.

Full cooperation may still be impossible under a performance-based firing rule if the lay-offs are sufficiently large. Propositions 2 suggests that the equilibrium with the most cooperation will then have conditional cooperators cooperate and egoists defect in some period $t \leq T_0$. In this case, moving from a random firing rule to a performance-based rule would not unambiguously increase equilibrium cooperation, because it may cause such an equilibrium to break down.

Proposition 7. *Suppose that a fraction $1 - \delta_0$ of players are laid off after period T_0 , in decreasing order of the belief that they are egoists: if $\mu_{iT_0+1} = \mu_{jT_0+1}$, then players i and j are equally likely to be laid off, and if $\mu_{jT_0+1} < \mu_{iT_0+1}$ and player i is laid off, then player j will be laid off as well. Suppose furthermore that $\delta_0 < \lambda$.*

Consider any candidate equilibrium where different types take different actions in some period $t \leq T_0$.

Then an egoist's incentive to make any deviation from this candidate equilibrium is now strictly larger than under a random firing rule.

A performance-based firing rule makes it unattractive for an egoist to reveal his type. If this effect is sufficiently strong, then the rule can help sustain an equilibrium where players always cooperate. Otherwise the rule will just tighten an egoist's incentive constraint in any candidate equilibrium with partial cooperation, including any sorting equilibrium, by pushing an egoist to imitate a conditional cooperator.

When lay-offs are large, $\delta_0 < \lambda$, an egoist who plays his equilibrium strategy is sure to be laid off. He then loses the discounted stream of pay-offs he would have received after the lay-offs had he remained in the game under a random firing rule. This pay-off of a per period is at least as high as he would receive after making any deviation, so the performance-based rule makes a deviation more attractive for any out-of-equilibrium beliefs.

This result raises questions related to a firm's firing policy and to issues of commitment. In a setting where workers differ only in their intrinsic motivation, laying off egoists could help safeguard future cooperation in the face of currently unforeseen shocks. The problem is that a performance-based firing rule can prevent players from revealing their type. It could even leave defection in all periods before the lay-offs as the unique equilibrium.

A manager could instead promise a different firing rule that does not penalize egoists, but this promise may not be credible. In the absence of commitment, a manager who prefers to lay off egoists would simply renege on his promise once he can identify them.

6 Robustness

Conditional cooperators can cooperate in multiple periods before the lay-offs in a sorting equilibrium because they are all able to rematch. Rematching may be difficult in practice, for example if workers require a transition period before they can work together productively. A natural question is then whether sorting can still promote cooperation if each player must remain with his initial teammate until period T_0 .

Not allowing for rematching will decrease the amount of cooperation in any sorting equilibrium, because a player who discovers his teammate is an egoist has nowhere to turn. The pair must still work together until period T_0 , and cooperation would break down whenever $\delta_0 < \delta_0^*$. Sorting would only allow a fraction λ of conditional cooperators to cooperate up until the lay-offs, those who learn that their teammate is also of the same type.

Although cooperation in any sorting equilibrium will decrease without rematching, the qualitative results

of the previous sections will remain unchanged. Propositions 1 and 2 describe when sorting can increase cooperation and both will continue to hold. Conditions (2) and (3) consider an egoist's incentive to deviate by defecting when all other players cooperate, which will trigger a punishment regardless of whether players can rematch. When either (2) or (3) is violated, a non-sorting equilibrium will have cooperation in no more than one period $t \leq T_0$, which is still less than in any sorting equilibrium without rematching.

The incentive constraints in a sorting equilibrium, (4) and (5) from Proposition 3, will change somewhat without rematching. A conditional cooperator will now have a lower incentive to sort, because the expected future benefit from revealing his type is a fraction λ of what it was with rematching. Looking at (4), a conditional cooperator's incentive constraint becomes

$$\lambda b + (1 - \lambda)d - \lambda(a + \Gamma) - (1 - \lambda)c \leq \lambda \left(\frac{\delta}{1 - \delta} (1 - \delta^{T_0 - t'}) (a + \Gamma - d) \right). \quad (7)$$

An egoist will now have a higher incentive to sort, because the expected future benefit from deviating is also reduced to a fraction λ of what it was with rematching. This is the probability that an egoist discovers his teammate is a conditional cooperator who he can later take advantage of. Looking at (5), an egoist's incentive constraint becomes

$$\lambda \left(\frac{\delta}{1 - \delta} (1 - \delta^{T_0 - t' - 1}) (a - d) + \delta^{T_0 - t'} (b - d) - \delta_0 \frac{\delta^{T_0 - t' + 1}}{1 - \delta} (a - d) \right) \leq \lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c. \quad (8)$$

The only difference between (4) and (7), and between (5) and (8), is the extra term λ that now multiplies one side of each expression. Comparative statics of the type described in Section 4 will continue to hold without rematching.

I have also assumed that the timing of lay-offs is certain, whereas in reality workers might just foresee a time period where lay-offs are more likely. This time period could correspond to an expected announcement on firm performance or a discussion of restructuring.

A way to capture this idea would be to assume that lay-offs could occur at most once over a stretch of N periods, beginning in period T_0 , where the conditional probability of lay-offs after any given period is $p \leq 1$. Setting $p = 1$ would then yield the original model.

Sorting would continue to play a role under uncertainty if full cooperation cannot be sustained prior to period T_0 . It is straightforward to show this will remain the case when $\delta_0 < \delta_0^*$, as long as p exceeds a certain threshold. The assumption that lay-offs are certain is not necessary for some cooperation to break down, it is enough that lay-offs are sufficiently likely. The threshold value is also decreasing in N . The longer the

period of uncertainty, the lower the value of p needed for egoists to defect in period T_0 and for sorting to play a role.

7 Conclusion

This paper has examined how sorting can help sustain cooperation in teams where some players are intrinsically motivated conditional cooperators, others are selfish egoists, and where lay-offs occur at a fixed future date. Unless players are sufficiently patient or the fraction of conditional cooperators is sufficiently high, then the equilibrium with the most cooperation will be a sorting equilibrium. Sorting allows players to reveal their type, so that conditional cooperators can identify each other and cooperate leading up to the lay-offs.

Certain changes that would seem to make cooperation more attractive may decrease equilibrium cooperation if they increase an egoist's incentive to sort. An increase in the discount factor, in the fraction of conditional cooperators, or a performance-based firing rule that targets selfish players can all push an egoist to deviate from a sorting equilibrium by imitating a conditional cooperator. Cooperation may then break down because conditional cooperators cannot reveal their type.

Appendix

Proof of Proposition 1. Consider a candidate equilibrium where both types cooperate in all periods. A conditional cooperator has no incentive to deviate because he earns the maximum pay-off $a + \Gamma$ in each period. The equilibrium pay-off for an egoist is

$$\sum_{t=1}^{T_0} \delta^{t-1} a + \delta_0 \frac{\delta^{T_0}}{1-\delta} a,$$

where δ_0 is the probability of remaining in the game after period T_0 . An egoist who deviates in some period $t' \geq T_0 + 1$ earns b but is then punished by the trigger strategy. The deviation will not be profitable if

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

This inequality is equivalent to $\delta < (b-a)/(b-d)$, which holds by (1). An egoist who deviates in some $t' \leq T_0$ earns

$$\sum_{t=1}^{t'-1} \delta^{t-1} a + \delta^{t'-1} b + \sum_{t=t'+1}^{T_0} \delta^{t-1} d + \delta_0 \frac{\delta^{T_0}}{1-\delta} d.$$

This pay-off is increasing in t' , since (1) implies $a + \delta b \geq b + \delta d$. The most profitable deviation has $t' = T_0$, and an egoist's incentive constraint becomes

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

This constraint is just

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

so the deviation is unprofitable whenever $\delta_0 \in [\delta_0^*, 1]$.

Proof of Proposition 2. Full cooperation is not possible by $\delta < \delta_0^*$, so consider a candidate equilibrium where both types cooperate in all periods except for egoists who defect in period T_0 .

A conditional cooperator will not deviate in any $t \leq T_0 - 1$, because he earns the maximum pay-off $a + \Gamma$. His pay-off as of period T_0 , discounted as of that period, is

$$\lambda(a + \Gamma) + (1 - \lambda)c + \delta_0 \frac{\delta}{1 - \delta} (a + \Gamma).$$

The first two terms give the pay-off from cooperating in period T_0 , where the probability his teammate is also a conditional cooperator is the prior λ . A deviation in period T_0 will not be punished because defecting mimics an egoist's equilibrium action. The subsequent pay-off from deviating is

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

The deviation will not be profitable if

$$\lambda b + (1 - \lambda)d \leq \lambda(a + \Gamma) + (1 - \lambda)c,$$

which is equivalent to (2). The pay-off for an egoist in this candidate equilibrium is

$$\sum_{t=1}^{T_0-1} \delta^{t-1} a + \delta^{T_0-1} (\lambda b + (1 - \lambda)d) + \delta_0 \frac{\delta^{T_0}}{1 - \delta} a.$$

A deviation in period T_0 cannot be profitable, since defection is an egoist's dominant strategy in the stage game. The only profitable deviation can be to defect in some period $t' \leq T_0 - 1$, which yields

$$\sum_{t=1}^{t'-1} \delta^{t-1} a + \delta^{t'-1} b + \sum_{t=t'+1}^{T_0} \delta^{t-1} d + \delta_0 \frac{\delta^{T_0}}{1 - \delta} d.$$

This pay-off is increasing in t' , since (1) implies $a + \delta b \geq b + \delta d$. The most profitable deviation has $t' = T_0 - 1$, which gives

$$\sum_{t=1}^{T_0-2} \delta^{t-1} a + \delta^{T_0-2} b + \delta^{T_0-1} d + \delta_0 \frac{\delta^{T_0}}{1-\delta} d.$$

The deviation will not be profitable if

$$b + \delta d + \delta_0 \frac{\delta^2}{1-\delta} d \leq a + \delta(\lambda b + (1-\lambda)d) + \delta_0 \frac{\delta^2}{1-\delta} a,$$

which is equivalent to (3).

An equilibrium cannot exist where both types cooperate in all periods, except for conditional cooperators who defect in period T_0 . An egoist could then deviate by defecting in period T_0 , increase his immediate pay-off from $\lambda a + (1-\lambda)c$ to $\lambda b + (1-\lambda)d$, and avoid any punishment because he mimics a conditional cooperator. When (2) and (3) both hold, the equilibrium with the most cooperation therefore has both types cooperate in all periods except for egoists who defect in period T_0 .

If (3) is violated but (2) is not, then another non-sorting equilibrium exists where both types defect for all $t \leq T_0$ except for conditional cooperators who cooperate in period T_0 . A deviation in any period $t' \leq T_0 - 1$ cannot be profitable, because it reduces a player's immediate pay-off from d to c and triggers a punishment. An egoist will not deviate in period T_0 because defection is his dominant strategy in the stage game, while a conditional cooperator will not deviate in period T_0 because (2) holds.

An equilibrium always exists where both types defect for all $t \leq T_0$, since deviating reduces a player's immediate pay-off and triggers a punishment. I now show that when (3) is violated but (2) is not, this is the unique non-sorting equilibrium. In any other non-sorting equilibrium, both types must cooperate in some period $t' \leq T_0 - 1$. An egoist who deviates in period t' will be punished in all later periods. Both types take the same action for any $t' + 1 \leq t \leq T_0 - 1$ in a non-sorting equilibrium, and the punishment only reduce an egoist's pay-off in period t if both types would have cooperated. The incentive to deviate is then lowest when both types cooperate for all $t \leq T_0 - 1$. This deviation is still profitable because (3) is violated.

If a sorting equilibrium exists, then there is some period $t' \leq T_0 - 1$ where different types first take different actions. By Bayes' rule, players then update their beliefs to $\mu_{it'+1} = 1$ if player i is a conditional cooperator and $\mu_{it'+1} = 0$ if he is an egoist. The matching rule pairs conditional cooperators together so they can cooperate for all $t \geq t' + 1$. Since $t' \leq T_0 - 1$, any sorting equilibrium involves more cooperation than a non-sorting equilibrium where both types defect for all $t \leq T_0 - 1$.

Proof of Proposition 3. Consider a candidate sorting equilibrium where different types first take different

actions in period $t' \leq T_0 - 1$. If both types defect for all $t \leq t' - 1$, then no player wants to deviate in any of these periods. The deviation would just reduce a player's immediate pay-off from d to c and trigger a punishment.

Egoists cannot cooperate and conditional cooperators defect in period t' , since then an egoist would prefer to deviate. By defecting in period t' , he could increase his immediate pay-off from $\lambda c + (1 - \lambda)a$ to $\lambda d + (1 - \lambda)b$, and earn a in all subsequent periods by mimicking a conditional cooperator. In the period where players sort, it follows that conditional cooperators must cooperate and egoists defect.

A conditional cooperator's pay-off as of period t' in a sorting equilibrium is

$$\lambda(a + \Gamma) + (1 - \lambda)c + \sum_{t=1}^{T_0-t'} \delta^t(a + \Gamma) + \delta_0 \frac{\delta^{T_0-t'+1}}{1 - \delta}(a + \Gamma), \quad (9)$$

and an egoist's pay-off as of period t' is

$$\lambda b + (1 - \lambda)d + \sum_{t=1}^{T_0-t'} \delta^t d + \delta_0 \frac{\delta^{T_0-t'+1}}{1 - \delta} a. \quad (10)$$

The probability that a player's teammate will cooperate in period t' is just the prior λ . By Bayes' rule, players then update their beliefs after period t' to $\mu_{it'+1} = 1$ if player i is a conditional cooperator and $\mu_{it'+1} = 0$ if he is an egoist. The matching rule pairs conditional cooperators together so they can cooperate and earn a in all later periods, while egoists must defect for all $t' + 1 \leq t \leq T_0$ by backwards induction.

It is sufficient to consider players' incentive to deviate by changing their period t' action. A conditional cooperator will not deviate in any $t \geq t' + 1$, because he earns the maximum pay-off $a + \Gamma$, while an egoist's deviation in any $t \geq t' + 1$ would reduce his immediate pay-off to c and trigger a punishment.

A conditional cooperator who deviates by defecting in period t' will mimic an egoist in all later periods, since an unexpected action would reduce his immediate pay-off and trigger a punishment. The pay-off from deviating is given by (10), but with a replaced by $a + \Gamma$.

Comparing (9) with (10), a conditional cooperator's incentive constraint is

$$\lambda b + (1 - \lambda)d + \sum_{t=1}^{T_0-t'} \delta^t d \leq \lambda(a + \Gamma) + (1 - \lambda)c + \sum_{t=1}^{T_0-t'} \delta^t(a + \Gamma),$$

which is equivalent to (4).

An egoist who deviates by cooperating in period t' earns $\lambda a + (1 - \lambda)c$ and rematches with a conditional cooperator. His teammate will cooperate as long as he continues to mimic a conditional cooperator and cooperate himself. The egoist's incentive to defect in any period $t \geq t' + 1$ will therefore be the same as in a candidate equilibrium where both types always cooperate. By the same argument as in the proof

of Proposition 1, (1) and $\delta_0 < \delta_0^*$ imply that an egoist who deviates in period t' will cooperate for all $t' \leq t \leq T_0 - 1$ and then defect in period T_0 .

An egoist's pay-off as of period t' from this deviation is

$$\lambda a + (1 - \lambda)c + \sum_{t=1}^{T_0-t'-1} \delta^t a + \delta^{T_0-t'} b + \delta_0 \frac{\delta^{T_0-t'+1}}{1 - \delta} d. \quad (11)$$

Comparing (10) with (11), an egoist's incentive constraint is

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

which is equivalent to (5). Rearranging (5) gives

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

where the left-hand side does not depend on Γ . Setting the critical value K_2 equal to this left-hand side, a sufficient condition for (5) is that $d - c \geq K_2$.

Recall that (4) is given by

$$\lambda b + (1 - \lambda)d - \lambda(a + \Gamma) - (1 - \lambda)c \leq \frac{\delta}{1 - \delta} (1 - \delta^{T_0-t'}) (a + \Gamma - d),$$

where the right-hand side is increasing without bound in Γ . For a given value of K_2 , define K_1 as the value of Γ for which (4) holds with equality. Then a sufficient condition for (4) is that $\Gamma \geq K_1$.

Proof of Proposition 4. If both (4) and (5) hold for some t' , then a sorting equilibrium will exist where both types defect for all $t \leq t' - 1$ and then conditional cooperators begin to cooperate. Consider another candidate sorting equilibrium where both types cooperate in some period $t'' \leq t' - 1$. For this to be an equilibrium, an egoist must not want to deviate by defecting in period t'' .

An egoist who defects in period t'' is punished in all later periods. This punishment only reduces his pay-off in periods where he would obtain more than d in the candidate equilibrium, so the incentive to deviate is lowest when the candidate equilibrium has both types cooperate for all $t \leq t' - 1$. An egoist's pay-off in this candidate equilibrium is

$$\sum_{t=1}^{t'-1} \delta^{t-1} a + \delta^{t'-1} (\lambda b + (1 - \lambda)d) + \sum_{t=t'+1}^{T_0} \delta^{t-1} d + \delta_0 \frac{\delta^{T_0}}{1 - \delta} a, \quad (12)$$

while his pay-off from deviating in period t'' is

$$\sum_{t=1}^{t''-1} \delta^{t-1} a + \delta^{t''-1} b + \sum_{t=t''+1}^{T_0} \delta^{t-1} d + \delta_0 \frac{\delta^{T_0}}{1-\delta} d.$$

This pay-off is increasing in t'' because (1) implies $a + \delta b \geq b + \delta d$. The most profitable deviation has $t'' = t' - 1$, which yields

$$\sum_{t=1}^{t'-2} \delta^{t-1} a + \delta^{t'-2} b + \sum_{t=t'}^{T_0} \delta^{t-1} d + \delta_0 \frac{\delta^{T_0}}{1-\delta} d. \quad (13)$$

An egoist will not deviate if (12) exceeds (13), which is equivalent to condition (6). When (6) holds, a sorting equilibrium will exist where conditional cooperators cooperate for all $t \leq T_0$, and egoists cooperate for $t \leq t' - 1$ and defect for $t' \leq t \leq T_0$.

The right-hand side of (6) is decreasing in t' , and so is easiest to satisfy for $t' = t'_H$. The sorting equilibrium with the most cooperation then has both types cooperate until $t' = t'_H$, then egoists begin to defect. If t'_H does not satisfy (6), then players cannot cooperate before period t' in any sorting equilibrium. The most cooperation occurs when players sort as early as possible, in period $t' = t'_L$.

Proof of Lemma 1. To establish the first part of the lemma, it is sufficient to show that the right-hand side of (4) and the left-hand side of (5) are both decreasing in t' . This is the case for (4) because $1 - \delta^{T_0-t'}$ is decreasing in t' . Adding and subtracting $\delta^{T_0-t'}(a - d)$ to the left-hand side of (5) gives

$$\frac{\delta}{1-\delta}(1 - \delta^{T_0-t'})(a - d) + \delta^{T_0-t'}(b - a) - \delta_0 \frac{\delta^{T_0-t'+1}}{1-\delta}(a - d). \quad (14)$$

Subtracting (14) from the same expression evaluated at $t' + 1$ yields

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

Dividing through by $\delta^{T_0-t'-1} - \delta^{T_0-1} > 0$ and rearranging shows this will be negative if

$$b - a \leq \frac{\delta}{1-\delta}(a - d)(1 + \delta_0). \quad (15)$$

The right-hand side of (15) is minimized when $\delta_0 = 0$ and $\delta = (b - a)/(b - d)$, so that (1) holds with equality. Then (15) also holds with equality, so the left-hand side of (5) is decreasing in t' .

The left-hand side of (5) is given by (14), and rewriting (4) gives

$$\lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c \leq \frac{\delta}{1 - \delta}(1 - \delta^{T_0 - t'})(a + \Gamma - d) + \lambda\Gamma.$$

To establish the second part, it is sufficient to show that

$$\frac{\delta}{1 - \delta}(1 - \delta^{T_0 - t'})(a - d) + \delta^{T_0 - t'}(b - a) - \delta_0 \frac{\delta^{T_0 - t' + 1}}{1 - \delta}(a - d) < \frac{\delta}{1 - \delta}(1 - \delta^{T_0 - t'})(a + \Gamma - d) + \lambda\Gamma. \quad (16)$$

The left-hand side of (16) is decreasing in δ_0 , while the right-hand side is increasing in Γ . I set $\delta_0 = 0$ and let Γ tend to $b - a$ to give

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

Simplifying shows this is equivalent to

$$\left(\frac{\delta^{T_0 - t'} - \delta}{1 - \delta} - \lambda \right) (b - a) < 0,$$

which holds because $\delta^{T_0 - t'} - \delta < 0$.

Proof of Proposition 5. Proposition 1 showed that players can cooperate in all periods if $\delta_0 \geq \delta_0^*$, which is equivalent to

$$\delta \leq \frac{b - a}{b - a + \delta_0(a - d)}. \quad (17)$$

Define δ_3 as the right-hand side of (17), where $\delta = \delta_3$ implies $\delta_0 = \delta_0^*$ and where $\delta_3 < 1$ by $\delta_0 > 0$. An equilibrium exists where both types cooperate in all periods if $\delta \in [\delta_3, 1]$.

Since $b > a$, $d > c$ and $\lambda < 1$, condition (2) will be violated for Γ close enough to $b - a$. Whenever $\delta < \delta_3$, Proposition 2 then shows that the only non-sorting equilibrium has both types defect in all $t \leq T_0$.

I now show that no sorting equilibrium exists when $\delta = \delta_3$. Condition (5) evaluated at $t' = T_0 - 1$ is

$$\delta(b - d) - \delta_0 \frac{\delta^2}{1 - \delta}(a - d) \leq \lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c, \quad (18)$$

while $\delta_0^*(a - d)[\delta/(1 - \delta)] = b - a$ means that (18) evaluated at $\delta = \delta_3$ is

$$\delta_3(b - d) - \delta_3(b - a) \leq \lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c. \quad (19)$$

Rearranging (19) gives

$$\delta_3(a - d) \leq \lambda(b - a) + (1 - \lambda)(d - c),$$

which I now show does not hold. Substituting for δ_3 from (17) yields

$$\left[\frac{b - a}{b - a + \delta_0(a - d)} \right] (a - d) \leq \lambda(b - a) + (1 - \lambda)(d - c).$$

Since $b - a > d - c$, it is enough to show that

$$\left[\frac{1}{b - a + \delta_0(a - d)} \right] (a - d) > 1,$$

which is satisfied by $b - a < (1 - \delta_0)(a - d)$. Lemma 1 then implies that (5) is also violated for all $t' \leq T_0 - 2$, and no sorting equilibrium exists for $\delta = \delta_3$.

The right-hand side of (18) is strictly positive, so (18) must hold when δ is small. Since the left-hand side of (18) is concave in δ , there is a unique value of $\delta \in (0, \delta_3)$ such that (18) holds with equality. Define δ_2 as this value of δ . No sorting equilibrium exists when $\delta \in (\delta_2, \delta_3)$, so the unique equilibrium has both types defect for all $t \leq T_0$.

Since (5) holds with equality at $t' = T_0 - 1$ and $\delta = \delta_2$, Lemma 1 implies that (4) must strictly hold. Moreover, (18) also holds for all $\delta \leq \delta_2$. By continuity, there must be some $\delta_1 < \delta_2$ such that a sorting equilibrium exists with $t' = T_0 - 1$, whenever $\delta \in [\delta_1, \delta_2]$.

To complete the proof, I verify that (1) holds for all $\delta \geq \delta_1$. It will be sufficient to show that (18) holds strictly when evaluated at $\delta = (b - a)/(b - d)$. Substituting and rearranging gives

$$(1 - \lambda)(b - a) - (1 - \lambda)(d - c) \leq \delta_0 \frac{\delta^2}{1 - \delta} (a - d),$$

which holds because the right-hand side is strictly positive and $(b - a) - (d - c)$ is small.

Proof of Proposition 6. Define λ_3 as the value of λ for which (2) holds with equality,

$$\lambda_3 b + (1 - \lambda_3)d - \lambda_3 a - (1 - \lambda_3)c = \lambda_3 \Gamma. \quad (20)$$

For all $\lambda \in [\lambda_3, 1]$, Proposition 2 shows that a non-sorting equilibrium exists where conditional cooperators cooperate in period T_0 . When λ is sufficiently close to 1, so that (3) also holds, a non-sorting equilibrium exists where both types cooperate in all periods except for egoists who defect in period T_0 .

I now show that no sorting equilibrium exists when $\lambda \in [\lambda_3, 1]$, because (5) is violated for all t' . First, let $t' = T_0 - 1$. Since $\delta = 1$ and $\delta_0 = 0$, (5) becomes

$$b - d \leq \lambda b + (1 - \lambda)d - \lambda a - (1 - \lambda)c. \quad (21)$$

When $\lambda = \lambda_3$, substituting from (20) into (21) yields

$$b - d \leq \lambda_3 \Gamma.$$

This inequality is violated for $a + \Gamma$ sufficiently close to b , because $b - d > b - a$. The right-hand side of (21) is decreasing in λ , as $b - d < d - c$ implies $b - d < a - c$. This means that (21) is also violated for all $\lambda \in [\lambda_3, 1]$ and a sorting equilibrium does not exist with $t' = T_0' - 1$. By Lemma 1, (5) is also violated for all $t' \leq T_0 - 2$, and no sorting equilibrium exists for any value of t' .

Since pay-offs are continuous in λ , (5) is violated for all t' whenever λ is marginally less than λ_3 . This implies there is an interval (λ_2, λ_3) where no sorting equilibrium exists. Condition (2) is also violated for all $\lambda < \lambda_3$, so the unique equilibrium when $\delta \in (\lambda_2, \lambda_3)$ has both types to defect in all $t \leq T_0$.

Condition (21) will hold when λ is small by $b - d < d - c$, so there is some λ_2 for which (21) holds with equality, and where (21) strictly holds for all $\lambda < \lambda_2$. When $\lambda = \lambda_2$, Lemma 1 implies that (4) must strictly hold for $t' = T_0 - 1$. This means that for $\lambda = \lambda_2$, an equilibrium exists where players sort in period $T_0 - 1$. Pay-offs are continuous in λ , so (5) also holds for λ marginally less than λ_2 , and there is an interval $\lambda \in [\lambda_1, \lambda_2]$ where this sorting equilibrium exists.

Proof of Proposition 7. Consider a candidate equilibrium where different types take different actions in some period $t \leq T_0$. Egoists reveal their type, so $\delta_0 < \lambda$ means that all egoists and $(1 - \delta_0) - (1 - \lambda)$ conditional cooperators are laid off. Compared to a random firing rule, an egoist's expected pay-off as of period $T_0 + 1$ decreases from $\delta_0 \frac{\delta}{1 - \delta} a$ to zero.

If an egoist deviates by imitating a conditional cooperator in all periods, he will be laid off with probability $1 - \delta_0/\lambda$. Compared to a random firing rule, his expected pay-off as of period $T_0 + 1$ increases from $\delta_0 \frac{\delta}{1 - \delta} a$ to $(\delta_0/\lambda) \frac{\delta}{1 - \delta} a$.

If an egoist deviates by taking an unexpected action, at worst he will be laid off with probability one. His expected pay-off as of period $T_0 + 1$ then decreases from $\delta_0 \frac{\delta}{1 - \delta} d$ to zero. Compared to a random firing rule, his expected pay-off from this deviation cannot decrease by more than $\delta_0 \frac{\delta}{1 - \delta} d < \delta_0 \frac{\delta}{1 - \delta} a$. This means that an egoist's incentive to make any deviation has strictly increased.

References

- Barker, J. R. (1993). Tightening the iron cage: Concertive control in self-managing teams. *Administrative Science Quarterly*, 38(3):408–437.
- Bartling, B. and von Siemens, F. A. (2010a). Equal sharing rules in partnerships. *Journal of Institutional and Theoretical Economics*, 166(2):299–320.
- Bartling, B. and von Siemens, F. A. (2010b). The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics*, 17(3):598–607.
- Bewley, T. F. (1999). *Why Wages Don't Fall during a Recession*. Harvard University Press, third edition.
- Burlando, R. and Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics*, 8(1):35–54.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83.
- Conlon, J. R. (2003). Hope springs eternal: learning and the stability of cooperation in short horizon repeated games. *Journal of Economic Theory*, 112(1):35–65.
- de Oleveira, A. C., Croson, R., and Eckel, C. (2009). Theories of fairness and reciprocity - evidence and economic applications. Working paper, Department of Resource Economics, University of Massachusetts-Amherst.
- Delfgaauw, J. and Dur, R. (2007). Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, 62(4):605–624.
- Delfgaauw, J. and Dur, R. (2008). Incentives and workers' motivation in the public sector. *Economic Journal*, 118(525):171–191.
- Delfgaauw, J. and Dur, R. (2009). From public monopsony to competitive market: more efficiency but higher prices. *Oxford Economic Papers*, 61(3):586–602.
- Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(35):556–590.
- Fehr, E. and Schmidt, K. M. (2001). Theories of fairness and reciprocity - evidence and economic applications. CEPR Discussion Papers 2703, C.E.P.R. Discussion Papers.

- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.
- Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing ‘conditional cooperation’ in a field experiment. *American Economic Review*, 94(5):1717–1722.
- Gächter, S. (2006). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. Discussion Papers 2006-03, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.
- Gächter, S. and Thoni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2-3):303–314.
- Gunnthorsdottir, A., Houser, D., and McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization*, 62(2):304–315.
- Keser, C. and van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, 102(1):23–39.
- Kosfeld, M. and von Siemens, F. A. (2009). Worker self-selection and the profits from cooperation. *Journal of the European Economic Association*, 7(2-3):573–582.
- Kosfeld, M. and von Siemens, F. A. (2011). Competition, cooperation, and corporate culture. *The RAND Journal of Economics*, 42(1):23–43.
- Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners’ dilemma. *Journal of Economic Theory*, 27(2):245–252.
- Lawler, E. and Cohen, S. (1992). Designing pay systems for teams. *American Compensation Association Journal*, 1(1):6–19.
- Lawler, E. A., Mohrman, S. A., and Benson, G. (2001). *Organizing for high performance: Employee Involvement, TQM, Reengineering, and Knowledge Management in the Fortune 1000*. Jossey-Bass, San Francisco.
- Lawler, E. A., Mohrman, S. A., and Ledford, G. E. (1995). *Creating high performance organizations*. Jossey-Bass, San Francisco.
- Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist’s view of human resources. *Journal of Economic Perspectives*, 21(4):91–114.

von Siemens, F. A. (2011). Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers*, 63(3):499–522.