Universidad de Navarra

Facultad de Ciencias Económicas y Empresariales

Working Paper nº 12/09

Peer Evaluations and Team Performance: When Friends Do Worse Than Strangers

Brice Corgnet
Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Peer Evaluations and Team Performance: When Friends Do Worse Than Strangers

Brice Corgnet

Working Paper No.12/09
November 2009

ABSTRACT

We use peer assessments as a tool to allocate joint profits in a real effort team experiment. We find that using this incentive mechanism reduces team performance. More specifically, we show that teams composed of fellows rather than strangers actually underperform in a context of peer evaluations. We conjecture that peer evaluations undermine the inherently high level of intrinsic motivation that characterizes teams composed of friends. We finally analyze the determinants of peer assessments and stress the crucial importance of equality concerns.

Brice Corgnet
Universidad de Navarra
Depto. Economía
Campus Universitario
31080 Pamplona
bcorgnet@unav.es

Peer Evaluations and Team Performance: When Friends Do Worse Than Strangers

October 26, 2009

Abstract

We use peer assessments as a tool to allocate joint profits in a real effort team experiment. We find that using this incentive mechanism reduces team performance. More specifically, we show that teams composed of fellows rather than strangers actually underperform in a context of peer evaluations. We conjecture that peer evaluations undermine the inherently high level of intrinsic motivation that characterizes teams composed of friends. We finally analyze the determinants of peer assessments and stress the crucial importance of equality concerns.

Keywords: team incentives, peer evaluations, experiments

JEL Classification Numbers: C92, M12, M54

1 Introduction

1.1 Team incentives and peer evaluations

In the last twenty years, more intensive use of teamwork in organizations has aroused interest with regard to the factors affecting the success of teams

(Lawler, Mohrman and Ledford 1995, Devine et al. 1999, Kozlowski et al. 1999). One of the crucial elements in the success of teams relates to the choice of the sharing rule for the joint outcome. On the one hand, paying team workers according to their individual contributions instead of assigning them a fixed share of the joint outcome reduces free riding behaviors (Alchian and Demsetz 1972, Holmström 1982).[1] On the other hand, rewarding workers according to their individual performance is likely to be costly since individual contributions are typically difficult to evaluate. An example of imperfect measures of performance is the use of subjective assessments to reward and promote employees (Prendergast 1993, Baker, Gibbons and Murphy 1994, Levin 2003, MacLeod 2003, Rajan and Reichelstein 2006). In that particular case, not only imperfect observability of individual contributions but also favoritism, politicking or social norms may prevent subjects from assessing partners according to their true relative contribution.

In this paper we analyze the impact of using peer assessments on team performance in a laboratory experiment in which three subjects are randomly matched to work on a real effort task. Subjects had to complete a real team-task for which individuals' effort and abilities determined joint profits. This allowed us to introduce a dimension of merit in subjects' decisions over peer evaluations.

We focus on a small-team context in which the different partners are likely to observe each others' levels of effort and could, in principle, use this information to evaluate their team partners' relative contributions.[2] Also, we consider

[1]Evidence of free riding in teams when the joint output is equally shared among partners has been found in legal partnerships (Leibowitz and Tollison 1980) as well as in medical group practices (Newhouse 1973, Gaynor and Pauly 1990, Encinosa, Gaynor and Rebitzer 1991). At the experimental level, Nalbantian and Schotter (1997) as well as van Dijk, Sonnemans and van Winden (2001) have found evidence of free riding in groups when team profits are equally shared among subjects.

[2]The focus on small teams can also be motivated by stressing that the success of small

a case in which the total amount of money to be allocated among team members is fixed and partners have to determine the share of the joint outcome assigned to each individual using peer evaluations. In our setting, subjects are not allowed to allocate an amount of money greater than their team output. That is, the allocation rule for the joint outcome has to be budget-balanced. In the case of budget-balanced allocation rules, one can derive from Holmström (1982) that purely self-interested partners will have incentives to lie and undermine the achievements of their co-workers preventing group members from being rewarded according to their relative contribution.[3] As a result, even in a context in which team members are able to assess the contribution of their partners without errors they will have no incentives to tell the truth in equilibrium and peer evaluations will affect neither the motivation of team members nor team performance.

However, peer ratings have been used in many disciplines such as engineering, management and medical sciences as a mechanism to gather information about individual contributions (Van Rosendaal and Jennett 1992, Conway et al. 1993, Dochy, Segers and Sluijsmans 1999, Ramsey and Wenrich 1999, Thomas, Gebo and Hellmann 1999, Clark, Davies and Skeers 2005, Tu and Lu 2005) suggesting that reliable information not accessible to outsiders can be extracted from such assessments. Our objective is to provide an experimental analysis that helps us assess the impact of such practices on team performance. There are two possible effects at play here. First, if team members effectively report an accurate estimate of others' achievements, subjects will end up being paid according to a measure of their relative contribution to the joint outcome so that free riding behaviors may be reduced. Second, the use of peer evalua-

teams is a preliminary step for the successful growth of organizations (Weber 2006).

[3] A proof of this result is available upon request.

tions, by focusing team partners' attention on each others' contribution may increase peer pressure and reduce free riding behaviors in teams (Kandel and Lazear 1992, Falk and Ichino 2006). Evidently, peer evaluations are likely to be imperfect as they can be subject to the influence of team partners through politicking activities or be driven by social norms or pure self-interest. However, peer evaluations allow for the implementation of payment schemes that are based, even though imperfectly, on individuals' contributions without the need for external monitoring. In sum, peer ratings can be seen as an inexpensive mechanism to collect information about workers' contributions. In particular, our system of peer ratings does not require an additional task in order to identify possible free riders in teams as is the case in the mechanism proposed by Bartlett (1995). Bartlett actually proposes to select randomly one of the team partners in order to complete a final exam that will determine the grade of all the team partners.

We have to emphasize that our approach is based on the fact that the relative contribution of the team members are not observable by a third party so that piece rate payment schemes based on individual achievements cannot be implemented. We already know that individual piece-rate payments tend to increase the level of performance of employees as is found in Booth and Frank (1999), Lazear (2000) and Prendergast (1999).[4] We do not contemplate either the possibility of rewarding team partners according to their relative performance as in the case of tournaments in which the best performer receives a bonus payment (Lazear and Rosen 1981). There exists experimental evidence that tournaments incentives tend to increase team partners' levels of effort compared to equal rewards (Dickinson and Isaac 1998, van Dijk, Sonnemans and van Winden 2001, Irlenbusch and Ruchala 2008, Sutter 2008).

[4] These papers do not consider the case of piece-rate incentives in teams, however.

In the absence of piece-rate or tournaments contracts we propose to investigate experimentally the effect on team performance of an allocation mechanism based on peer evaluations. We show that peer evaluations undermine team performance when team members know each other. We account for this result by stressing that peer evaluations are likely to weaken the inherently high level of intrinsic motivation that is usually observed in teams formed by friends. This result gives support for the use of fixed sharing rules in partnerships and entrepreneurial teams that are frequently formed among friends.

1.2 Related experiments

In general, issues related to cooperation in teams can be analyzed experimentally using public good games. In particular, our benchmark design in which group members are rewarded a fixed share of the joint outcome can be seen as a one-shot Voluntary Contributions Mechanism (VCM) with the specificity that subjects work side by side on a real team task in which face to face communication is possible. In a VCM subjects decide upon the fraction of the initial endowment that they are willing to contribute to a group account, while keeping the remainder. The VCM is such that subjects maximizing their own monetary payoffs will not find it profitable to contribute to the group account while group earnings are maximized when everybody fully contributes to the collective account. The VCM literature stresses that contributions to the collective account increase when subjects are allowed to sanction or reward each other (Fehr and Gächter 2000, Masclet et al. 2003, Sefton, Shupp and Walker 2007) or when subjects are allowed to communicate (Isaac and Walker 1988, Sally 1995). In a comparison of these two mechanisms, Bochet, Page and Putterman (2006) showed that communication, especially when it is face to face, is a more effective mechanism than punishments in reducing free riding behaviors. As a result, we would expect a low level of free riding in our benchmark design since subjects are involved in a real team task in which face to face communication is possible. Regarding the use of peer evaluations, we can relate this payment scheme to VCM experiments in which subjects are given the possibility to reward or punish their partners (Sefton, Shupp and Walker 2007). However, as opposed to the results obtained in the VCM literature with punishments and rewards, the introduction of peer assessments in our setting tends to reduce group performance. This difference may simply result from the fact that subjects perceive sanctions and rewards in a VCM game very differently from peer assessments in our real team task experiment. In particular, Fehr and Rockenbach (2003) showed that in a trust game environment sanctions do not systematically increase cooperation if they are perceived as unfair. The authors stress that sanctions that are perceived as unfair destroy altruistic cooperation whereas sanctions perceived as fair tend to promote cooperation. In our experimental design, the introduction of a real team task leads subjects to consider the equal sharing of the joint outcome as the fair norm of behavior (Konow, Saijo and Akai 2008, Corgnet, Sutan and Veszteg 2009). As a result, peer assessments that deviate from the equal splitting norm may be seen as unfair and may reduce cooperation and team performance as a result.[5] In addition, the equality norm appears to be stronger in the case of classmates suggesting that the existence of interpersonal links may foster the negative perception of payment schemes based on peer evaluations . This sheds light on the fact that peer evaluations have a more negative effect on team performance when subjects know each other.

Our approach also relates to the experimental literature that stresses the

[5] See Section 3.4 for a discussion on the effect of peer evaluations on subjects' levels of satisfaction.

crowding-out effect of monitoring on the level of effort exerted by agents in standard principal-agent games. For example, Falk and Kosfeld (2006) implemented monitoring in a principal-agent game by giving the principal the possibility to impose a minimum level of effort on the agent. The authors found that most principals decided not to impose a minimum level of performance on the agent. This strategy led, on average, to a higher level performance compared to the monitoring strategy in which principals control the level of effort of the agent. This behavior of the principal can be explained by the presence of negative reciprocity that induces agents to lower their level of effort as a response to the signal of distrust created by the principal's decision to impose a minimum level of performance.

Frey (1993) argues that the crowding-out effect associated to monitoring activities is going to dominate its disciplining effect in the case of interpersonal relationships whereas the reverse would be true in the case of abstract relationships. In line with this analysis, we find that in teams composed of friends or classmates peer assessments tend to crowd-out motivation whereas this is not the case for teams that are composed of strangers. Our result is also consistent with Dickinson and Villeval (2008) in which they found some evidence of crowding-out of intrinsic motivation for intense levels of monitoring. This effect occurred in the case of non-anonymous subjects as long as distributional issues were at work. The authors also found extensive evidence supporting the disciplining effect of monitoring especially in the case in which the principal and the agent were strangers. The limited crowding-out effect encountered in Dickinson and Villeval (2008) may be due, as the authors themselves indicate, to the procedure used to create interpersonal relationships among subjects. In our design, interpersonal relationships exist because subjects know and see each other during the team task whereas in Dickinson and Villeval (2008) these interpersonal links are created by letting partners introduce themselves to each other at the beginning of the experimental session. The strength of the interpersonal relationship is then likely to be stronger in our design and this may explain why crowding-out of intrinsic motivation appears to be stronger in our case.

The rest of the paper is organized as follows. We present the design of the experiment in Section 2 while the results are introduced in the third section. Section 4 concludes and instructions are available in the appendix.

2 Design

2.1 Subject pool

We invited subjects through campus-wide posters and e-mail advertisements to participate in our experimental sessions at the University of Navarra in Spain. We recruited 66 subjects that were informed that the experiment concerned decision making and that it would last for approximately 60 minutes. Our experiments were run in sessions with twenty one or twenty four subjects depending on the treatment. Subjects were also informed that apart from a show-up fee of 5 euros they would receive a certain amount of money depending on their performance in the laboratory. The majority of students had either a Business or an Economics major and were completing the third year of their degree. The average earnings for the three experimental sessions were 20.79 euros. The details of the number of subjects and earnings for each of the three treatments are presented in the next table.

Table 1: Number of subjects and average earnings for the three treatments

Treatment	Subjects	Earnings
Equal split (**EQS**)	$n = 21$	21.70 €
Symmetric Peer evaluations (**SPE**)	$n = 24$	20.36 €
Asymmetric Peer evaluations (**APE**)	$n = 21$	20.36 €

2.2 Experimental procedure

The experiment was divided in two stages.

In the first stage of the game, subjects were randomly assigned to a team of three partners in order to complete a real effort task.[6] The team task is chosen so as to ensure that the following unique characteristics of teams were present: task interdependence, reward interdependence and social identity (Colquitt, Zapta-Phelan and Roberson 2005). Task interdependence was present in our design since some level of communication and coordination among subjects was required, for example, to avoid finding the same correct answers to the task. Also, reward interdependence followed from the payment scheme under which payoffs depended directly on the performance of the team. Finally, social identity could also develop since we had subjects sit side-by-side and participate in a real-effort task in which communication between team members was allowed. In addition, subjects were assigned the same function in the team so that no status was created.[7]

At the start of the experiment, each team member was assigned a letter

[6] Other real-effort tasks have been considered in the literature such as solving mazes (Gneezy, Niederle and Rustichini, 2003), completing sudokus (Calsamiglia, Franke and Rey Biel, 2009) or solving an optimization problem (Dickinson and Villeval 2008, van Dijk, Sonnemans and van Winden 2001, Montmarquette et al. 2004).

[7] Our approach can be seen as reward allocation between team members, partners or co-workers at the same layer of the hierarchy.

that indicates his position. Subject L was seated on the left of subject M whereas Subject R was placed on his right. Subjects had to find three- and four- digit numbers fulfilling certain conditions in 18 minutes as is described below. Each correct number was rewarded 45 euro cents while, each incorrect number incurred a penalty of 25 euro cents. Each team had access to only one set of instructions and to only one answer sheet so as to avoid subjects working separately.[8]

- Task 1 -

You have 18 minutes to find as many numbers as you can, satisfying the following conditions:

- It has 3 or 4 digits.
- If you sum its digits the result is equal to 15.
- If you multiply its digits the result is strictly larger than 10.
- The last two digits are strictly larger than 1.
- The first digit is an odd number.
- The second digit is an even number.

In the second stage of the game subjects were separated and had to answer a series of questions, individually and without the possibility of communication, in order to determine the allocation of the joint outcome obtained in the first stage. These questions as well as the splitting rule for the team outcome varied according to the treatment under study. Then, subjects had to complete individually Task 2 which was exactly the same exercise as Task 1 except that in the second condition the sum of the digits of the number had to give a result

[8] Each answer sheet consisted of a table of sixty-five cells where subjects could introduce a number in each cell.

equal to 14 instead of 15. As a result, the performance of subjects on Task 2 measures not only the individual ability of the different subjects but also their involvement in the team task since they were basically asked to repeat the team exercise alone. In Task 2, each correct number was rewarded 30 euro cents while, each incorrect number implied a penalty of 15 euro cents.[9]

The allocation rule for the team outcome was determined as follows for the three treatments.[10]

Equal split treatment (EQS) Before undertaking the team task in Stage 1, each team member knew that he would be rewarded one third of the team profits. In this treatment the allocation rule for the team outcome is exogenous while in the following two treatments the sharing rule for the joint outcome is determined by peer assessments.

Symmetric peer evaluations (SPE) Before undertaking the team task in Stage 1, each team member knew that the allocation rule for the team outcome would depend on his partners' evaluations. In particular, each subject $i \in \{L, M, R\}$ decides anonymously an allocation rule for the team output. That is, team partners choose a triple $\boldsymbol{\alpha}_i \equiv (\alpha_{iL}, \alpha_{iM}, \alpha_{iR})$ where α_{ij} is the percentage share of the team profits that subject i decides to assign to subject j, where $j \in \{L, M, R\}$ and $\alpha_{iI} + \alpha_{iL} + \alpha_{iM} = 1$.[11] Subjects knew that the share of the joint outcome obtained by a team partner would be computed as the average of the allocation rules chosen by the three team members. That

[9] We decided to reduce the earnings associated to each correct answer in the individual task compared to the team task. We did so in order to create synergies for teamwork. This particular choice does not affect our analysis.

[10] The decision about the allocation of joint profits was made before subjects knew the exact amount of money they earned on the task. Subjects could obviously form an accurate expectation of their earnings by multiplying the number of answers by 45 euro cents.

[11] The allocation rule is budget balanced.

is, the share of the joint profits obtained by subject j is computed as follows: $\frac{\alpha_{Lj}+\alpha_{Mj}+a_{Rj}}{3}$.

Under the standard assumption that subjects are self-interested we should expect that the symmetric peer evaluations treatment (**SPE**) does not affect team performance compared to the equal split treatment. Indeed, under this hypothesis, subjects will choose $\alpha_{ii} = 100\%$ for $i \in \{L, M, R\}$ so that each team partner finally receives one third of the team outcome as in the case of the equal split treatment. However, if we consider that subjects assess others' contributions taking into account their actual level of performance we should expect team partners' motivation and team performance to increase.

Asymmetric peer evaluations (APE) Before undertaking the team task in Stage 1, team members knew that the allocation rule for the team outcome would depend on their neighbors' assessments. In particular, subjects L and R decide anonymously which proportion of the team profits to assign to subject M. The share of the profits actually obtained by subject M is equal to the average of the sharing rules chosen by subjects L and R, that is $\frac{\alpha_{LM}+\alpha_{RM}}{2}$. At the same time, subject M decides, separately and anonymously, the proportion of the remaining profits to allocate to subjects L and R, respectively. The share of the team outcome assigned to subject L [R] is then computed as follows: $\left(1 - \frac{\alpha_{LM}+\alpha_{RM}}{2}\right)\alpha_{ML}$ $\left[\left(1 - \frac{\alpha_{LM}+\alpha_{RM}}{2}\right)\alpha_{MR}\right]$ where $\alpha_{ML} + \alpha_{MR} = 1$.

Treatment (**SPE**) is characterized by a complete system of peer evaluations in which each subject evaluates the contribution of his two team partners whereas, in treatment (**APE**), subjects evaluate only the contribution of their neighbor(s) as is described in Figure 1.

Under the standard assumption that subjects are purely self-interested we should expect that the asymmetric peer evaluations treatment (**APE**) increases

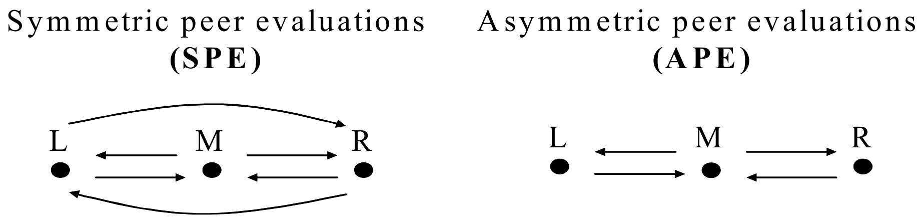

Figure 1: Peer evaluations treatments and the structure of assessments. Arrows indicate that a given subject assesses the contribution of the targeted team partner.

the level of motivation of subjects L and R compared to the equal split treatment whereas subject M should decide not to work on the task. Indeed, under this hypothesis, subjects L and R will set $\alpha_{LM} = \alpha_{RM} = 0$ so that subject M does not receive anything from joint profits. However, if we consider that team partners tell the truth about partners' achievements we may expect that free riding behaviors will be reduced as in the case of treatment (**SPE**). Also, if subjects L and R exhibit fairness concerns they may decide to share the team outcome equally among the three partners so that $\alpha_{LM} = \alpha_{RM} = \frac{1}{3}$.[12] If subject M anticipates such fair behavior he may decide to put effort into the task.

At the end of the experimental session subjects had to complete a questionnaire that was used to assess their level of motivation in the different tasks, their level of satisfaction in the tasks and their possible connections with the other team partners.

[12] There is evidence of fairness concerns in a large number of bargaining experiments such as the ultimatum game (Güth, Schmittberger and Schwarze 1982). In the ultimatum game, a subject (the proposer) receives an amount of money from the experimenter that he has to share between himself and a responder. The responder can reject the offer, in which case neither subject earns anything. If the responder accepts the division of the initial amount of money then it is implemented. In these experiments, close to fifty percent of the proposers offer an equal split of the team outcome.

3 Results

We present our results as well as the statistical analysis supporting our findings in the next subsections.

3.1 Payment schemes and team performance

We first compare team performance in the different treatments. Our findings are summarized as follows.

Result 1. (Equal split versus peer evaluations)

The use of peer evaluations leads to lower levels of performance than the equal split allocation rule.

Result 1 is illustrated by displaying the average team and individual performances in the three treatments.

Table 2: Average performance on the tasks
(Actual performance / Maximum performance)[13]

	Team task performance	Individual task performance
Equal split (**EQS**)	89.4%	51.3%
Symmetric Peer Evaluations (**SPE**)	79.1%	48.7%
Asymmetric Peer Evaluations (**APE**)	81.5%	42.5%
Combined (**SPE&APE**)	80.3%	45.8%

We stress that the team performance in the equal split treatment is 11.3% higher than in the peer evaluations treatments and this difference is statistically significant as can be seen in Table 3 below. The high level of team

[13]Maximum team task profits are equal to 29.7 euros and maximum profits on the individual task are equal to 15.9 euros.

performance in the equal split treatment is consistent with the study of van Dijk, Sonnemans and van Winden (2001) in which team incentives led to high levels of performance in a real effort public good game. Our result is also consistent with the fact that in VCM games very low levels of free riding are actually observed when face to face communication is possible as is the case in our experimental design (Sally 1995, Bochet, Page and Putterman 2006).

We report in Table 3 the p-values for Student tests and non-parametric Wilcoxon rank-sum tests. The alternative hypothesis is that the mean team performance or the mean individual performance is higher (or not equal) under the equal split treatment compared to peer evaluations treatments.

Table 3: Comparison of team performances across treatments

	Team task performance	Individual task performance
Treatment & Alternative hypothesis	P-values for *t-tests* (*Wilcoxon tests*)	P-values for *t-tests* (*Wilcoxon tests*)
EQS $>$ **SPE**	0.0264** (*0.0527**)	0.3187 (*0.3494*)
EQS $\neq$ **SPE**	0.0464** (*0.0928**)	0.6375 (*0.6987*)
EQS $>$ **SPE&APE**	0.0322** (*0.0959**)	0.1512 (*0.1753*)
EQS $\neq$ **SPE&APE**	0.0644* (*0.1918*)	0.3024 (*0.3506*)

We conclude that team performance is lower under the peer evaluations treatments. We are then in a case in which rigid sharing rules actually outperform a decentralized mechanism based on peer assessments. We consider a possible explanation for this puzzling result in the following subsection by focusing on the crucial role of friendship in mediating the effect of peer assessments on team performance.

3.2 Fellowship and team performance

We analyze in this section the mechanism underlying the effect of peer evaluations treatments on team performance by considering the role of fellowship among team members.

Result 2. (Fellowship and team performance)

Fellowship among team members tends to reduce team performance in the case of sharing rules based on peer evaluations.

Support for Result 2.

We estimate by OLS the following equation.[14]

$$P_i = \alpha_0 + \alpha_1 F_i + \alpha_2 X_i + \alpha_3 EQS_i + \alpha_4 EQS_i \times F_i + \varepsilon_t$$

where P_i is the joint profits of team i, X_i is an aggregate measure of the experience of subjects in team i in similar tasks, and EQS_i stands for a dummy variable that takes value one if team i has been involved in the equal split treatment (**EQS**) and takes value zero otherwise. We denote F_i a measure of the degree of fellowship in team i that accounts for the existing connections between the different partners of the team. The measure of fellowship is based on a questionnaire completed anonymously by subjects at the end of the experimental session. Subjects were asked to indicate whether each team partner was a stranger, a classmate or a friend. The measure of fellowship takes value $j \in \{0, 1, 2, 3\}$ if there exist exactly j different pairs of team partners that mutually acknowledge either knowing each other or being friends.[15] We display the distribution of teams according to fellowship levels in the following table.

[14] We also tried non-linear specifications for team partners' experience (X_i) that gave similar results.

[15] We identified three cases in which a subject claimed to know someone that actually claimed to be a stranger.

Table 4: The fellowship variable

Fellowship (F_i)	% of teams
$F_i = 0$	27.3%
$F_i = 1$	45.5%
$F_i = 2$	22.7%
$F_i = 3$	4.5%

Concerning the level of experience of team members, we use data on subjects' participation in related experiments in which subjects had to solve similar number tasks individually and by pairs (Corgnet, Sutan and Veszteg 2009). In this type of tasks there actually exists a learning effect so that people repeating the task usually do better than subjects completing this type of task for the first time. We control for this effect by introducing the variable X_i that measures the aggregate experience of subjects in team i. In particular, X_i takes value $v \in \{0, 1, 2, 3\}$ if v members of team i already participated in a similar experiment. We also build an aggregate measure of the ability of the team partners by considering either the average (AV) or the maximum performance (MAX) of the three partners on the individual task performed in Stage 2. We obtained similar regression results in those two cases as is shown in Table 5 below.[16] The regression estimates are displayed in the following table.

[16] We face an endogeneity issue when using the variables AV and MAX as estimates of subjects' individual abilities since it is likely that subjects working hard in the team task in Stage 1 will develop the specific skills that will help them reach a high score on the individual task in Stage 2.

Table 5: OLS regression

Dependent variable Team performance	Coefficient (Standard errors)	Coefficient (Standard errors)	Coefficient (Standard errors)
Fellowship (**F**)	-3.994^{***} (0.947)	-3.271^{***} (0.943)	-3.104^{***} (1.011)
Partners' experience (**X**)	1.283^{*} (0.712)	-	-
Partners' average ability (**AV**)	-	0.209 (0.125)	-
Partners' maximum ability (**MAX**)	-	-	0.094 (0.070)
Treatment (**EQS**)	-1.569 (1.644)	-1.809 (1.644)	-1.472 (1.692)
Interaction (**EQS&F**)	4.800^{***} (1.283)	3.826^{***} (1.232)	3.802^{***} (1.283)
R^2 (22 observations)	0.6071	0.5982	0.5772

As a result of our regression analysis in the case in which partners' experience is used as a control variable for group members' abilities, the slope for fellowship is positive for the equal split treatment ($-3.994 + 4.800 = 0.806$) whereas it is negative in the case of peer evaluations treatments (-3.994).[17] Our regression results are similar when either the average performance or the maximum performance of group members are used as controls for partners' abilities. We conclude that the negative impact of peer evaluations on team

[17] The coefficients associated to fellowship (α_1) and to the interaction effect (α_4) appeared to be highly significant in all the specifications of the model that we actually estimated. Notice that *** appears on a coefficient estimate if p-value < 0.01 and ** appears if $0.01 <$p-value < 0.05 whereas * appears if $0.05 <$p-value < 0.10.

performance is mediated by the magnitude of fellowship in teams. That is, peer assessments tend to reduce joint profits when subjects actually know each other. On the contrary, fellowship tends to facilitate cooperation and team performance when joint profits are shared equally. This result is interesting since it helps us understand the popularity of the equal sharing of joint outcomes in contexts in which fellowship is likely to play a role as in the case of partnerships and entrepreneurial teams.[18] We illustrate Result 2 in the following table by stressing that in the absence of fellowship team performances tend to be higher under peer evaluations whereas teams characterized by high levels of fellowship tend to perform better under an equal splitting rule.

Table 6: Fellowship and team performance

Average team performance	No fellowship $F_i = 0$	High fellowship $F_i \in \{2, 3\}$
Equal split (**EQS**)	89.9%	95.1%
Combined (**SPE & APE**)	94.1%	70.1%

We stress a possible interpretation of Result 2 based on the importance of intrinsic motivation in tasks performed among friends.[19] The use of peer ratings among fellows or friends may actually undermine subjects' motivation if team members are intrinsically motivated to contribute to the team and to

[18] The prevalence of the equal split sharing rule in entrepreneurial teams is stressed in the *Entrepreneurship and Compensation Survey* undertaken by Noam Wasserman in 2007.

[19] Other interpretations are related to negative reciprocity (see Section 3.4) and the importance of influence costs in teams in the presence of peer evaluations. Influence costs are defined as costs incurred to bias partners' decisions in one's own favor as well as the costs incurred to counter others' influence (Milgrom and Roberts 1992). We conjecture that influence costs are likely to arise in a context in which team partners know each other and have to decide about each others' rewards. However, we were not able to record subjects' discussions so as to assess possible attempts to influence team partners' evaluations.

help their partners. This is likely to be the case among friends that are inclined to behave altruistically to each other. Altruistic individuals are actually able to reduce free riding in teams as is exposed by Rotemberg (1994, 2006) since it permits subjects to commit on exerting high levels of efforts. However, in line with the self-perception theory developed in Social Psychology (Lepper 1973, Lepper, Greene and Nisbett 1973) the presence of peer evaluations may decrease friends' intrinsic motivation. Indeed, team partners that are naturally inclined to help each other may fear that in the presence of peer evaluations their altruistic behavior could be interpreted as purely self-interested. As a result, subjects may refrain from contributing to the joint effort. The fact that explicit rewards tend to obscure the true motives of helping behavior has been formalized in Bénabou and Tirole (2006). In order to provide support for this interpretation of Result 2 we decided to run an additional session for both the equal sharing treatment (**EQS**) and the symmetric peer evaluations treatment (**SPE**). In order to ensure that interpersonal relationships were sufficiently high we decided to run these sessions with classmates recruited from a course in Organization Behavior as is detailed in the next section.

3.3 Classmates, peer evaluations and team performance

We recruited 99 subjects that were informed that the experiment concerned decision making and that it would last for approximately 60 minutes. Our experiments were run in sessions with thirty or thirty nine subjects depending on the treatment. Subjects were also informed that apart from a show-up fee of 5 euros they would receive a certain amount of money depending on their performance in the laboratory. The students recruited for each treatment were classmates so that in a given group everybody knew each other, that is $F_i = 3$. Subjects had a Business major and were completing the third year of their degree. The average earnings for the three experimental sessions with classmates were 20.79 euros. The details of the number of subjects and earnings for each of the three treatments are presented in the next table.

Table 7: Number of subjects and average earnings
(Treatments with classmates)

Treatment	Subjects	Earnings
Equal split with classmates (**EQSC**)	$n = 30$	24.93 €
Symmetric Peer evaluations (classmates) (**SPEC**)	$n = 30$	18.97 €
Endogenous Team Formation (classmates) (**E-SPEC**)	$n = 39$	16.36 €

In treatment (**E-SPEC**) we consider a variation of the symmetric peer evaluations treatment in which subjects selected their team partners at their entrance to the laboratory so as to ensure that groups would be mainly composed of friends.[20] Indeed, the large majority of subjects recognized to use friendship as the main criterion to compose their group as is summarized in Table 8.

Table 8: Criteria for the selection of team partners

	Friendship	Cooperative Person	Socially Active	Talented Person	No Choice
Proportion of subjects	61.5%	21.5%	6.2%	1.6%	9.2%

We now compare team performance across the different treatments with classmates. As is shown in Table 9, team performance under the equal splitting rule is 20.4% higher than under peer evaluations. In the treatments that did not

[20] Subjects had 10 minutes to choose team partners in the room. If at the end of the 10 minutes some subjects were not matched the experimenter would form the remaining groups arbitrarily. It appears that all subjects were matched at the end of the 10 minutes period so that the experimenter did not have to intervene.

involve classmates (**EQS**, **SPE** and **APE**) this difference in team performance was only equal to 11.3%.

Table 9: Average performance on the team task

Treatment	Team task performance
Equal split (**EQSC**)	83.9%
Symmetric Peer evaluations (classmates) (**SPEC**)	69.7%
Endogenous Team Formation (**E-SPEC**)	64.2%[21]

We conjecture from the results displayed in Table 9 that when interpersonal relationships are important as is the case in teams composed of classmates the negative effect of peer evaluations on performance is particularly strong. We illustrate this conjecture in Figure 2 below in which we represent average team performance for the different treatments with and without classmates.

We provide statistical support for the fact that fellowship affects team performance negatively when peer evaluations are used in Table 10 below.

We conclude similarly to the case of subjects that were not classmates that team performance is on average significantly higher under the equal splitting rule compared to peer evaluations treatments. This finding provides additional support for Result 1. In addition we find that team performance in peer evaluations treatments is significantly lower in the groups composed of classmates in which interpersonal relationships are strong ($F_i = 3$) compared to groups that did not involve classmates. This finding confirms Result 2, that is the

[21]Team performance is especially low because of an atypicial observation with a team performance below 20%. We obtained similar results for the tests presented in Table 10 when this observation is dropped from the sample. In this case the average team performance in treatment **E-SPEC** is 68.5%.

Figure 2: Team performance across treatments, measured as the ratio between actual profits generated by the team in Task 1 and the maximum level of profits.

negative effect of peer evaluation on team performance increases when team partners know each other.

Table 10: Average performance on the team task and statistical tests

Treatment & Alternative hypothesis	P-values for *t-tests* (*Wilcoxon tests*)
EQSC $>$ **SPEC**	0.0400** (*0.0559**)
EQSC $\neq$ **SPEC**	0.0800* (*0.1119*)
EQSC $>$ **E-SPEC**	0.0226** (*0.0431***)
EQSC $\neq$ **E-SPEC**	0.0452** (*0.0863**)
EQSC $>$ **SPEC&E-SPEC**	0.0201** (*0.0279***)
EQSC $\neq$ **SPEC&E-SPEC**	0.0403** (*0.0559**)
APE&SPE $>$ **SPEC&E-SPEC**	0.0238** (*0.0609***)
APE&SPE $\neq$ **SPEC&E-SPEC**	0.0476** (*0.1219*)

3.4 Intrinsic motivation and the negative effect of peer evaluations on team performance

We investigate in this section the crowding-out of intrinsic motivation as a possible interpretation for the negative relationship that exists between fellowship and team performance when peer evaluations are used to allocate joint earnings. It is worth noting that both reciprocity effects and crowding-out of intrinsic motivation could potentially account for a reduction in effort and team performance in peer evaluations treatments. However, given our experimental design and in particular the fact that the game is not repeated, we decide to put emphasis on the crowding-out of intrinsic motivation. We have to stress that at no time in the experiment can subjects observe how their partners actually evaluated their contribution. This static structure of the game differs from standard principal-agent games evoked in the introduction (e.g. Falk and Kosfeld 2006) in which agents know the level of wages chosen by the principal before exerting an effort and principals know the agent's level of effort before setting wages in the next period. In our game, reciprocity is undermined by the static structure of the design and the fact that subjects are very unlikely to be able to assess with precision the level of monitoring of their team partners. We then focus on the crowding-out of intrinsic motivation as the main driving force behind our results.

To that end, we use subjects' answers to a questionnaire that they had to complete at the end of the experimental session in which they had to assess their level of motivation in the different tasks, their level of satisfaction in the tasks as well as their possible connections with the other team partners. Concerning the question on motivation, subjects involved in the peer evaluations treatments had to choose one of the following statements.

1- I felt motivated to do well in the team task because I knew my team partners would evaluate my work.

2- The fact that my team partners would evaluate my work did not affect my motivation on the team task.

The results of the questionnaire are displayed in the following table in which we include the results of our two initial sessions (**SPE** & **APE**) as well as the results of the two additional sessions with classmates (**SPEC** and **E-SPEC**).

Table 11: Motivation and peer evaluations[22]

Motivation in team Task 1 compared to individual Task 2	Not Increased	Increased
Symmetric Peer Evaluations (**SPE**) (24 subjects)	14	10 (41.7%)
Asymmetric Peer Evaluations (**APE**) (21 subjects)	12	9 (42.9%)
Symmetric Peer Evaluations (classmates) (**SPEC & E-SPEC**) (69 subjects)	51	18 (26.1%)

We conclude that teams composed of subjects who know each other (classmates) tend to be significantly less motivated by the use of peer evaluations than teams composed of strangers.[23] This finding is consistent with the interpretation of Result 2 based on the crowding-out hypothesis under which friends are likely to respond negatively to the use of a mechanism like peer evaluations

[22] We do not analyze separately teams for which the main criterion for group selection was not friendship since only three teams were not formed on the basis of this criterion.

[23] We reject the test that the proportion of subjects that did not feel motivated by the use of peer evaluations is the same for the two classmates samples and for the two initial samples (p-value = 0.07). The alternative hypothesis is that the proportion of subjects that did not feel motivated by the use of peer evaluations is higher for the two classmates samples than for the two initial samples.

Similar results are obtained if we include in the classmates sample all the teams from the two initial samples for which all partners knew at least one of the other team members (that is, $F_i > 1$).

that explicitly reward their contribution to the team. In order to provide support for the hypothesis that peer evaluations reduce intrinsic motivation among friends we compare subjects' levels of satisfaction across treatments. The level of satisfaction of subjects in the team task is assessed by the following question. How would you qualify your teamwork experience in Task 1?

1-Very poor, 2-Poor, 3-Acceptable, 4-Good, 5-Very Good.

We would expect that under the crowding-out of intrinsic motivation hypothesis classmates tend to exhibit a higher level of satisfaction than strangers in the case of equal splitting rules whereas this superior satisfaction should not be observed in the case of peer evaluations treatments. We would also expect that classmates exhibit a higher level of satisfaction in the equal sharing treatment compared to the peer evaluations treatment. The results provided in the next table are consistent with these predictions.

Table 12: Level of satisfaction across treatments

Treatment	**EQSC**	**SPEC**	**EQS**	**SPE&APE**
Mean satisfaction level	3.9	3.3	3.5	3.5

Table 13: Level of satisfaction across treatments and statistical tests

Treatment & alternative hypothesis	P-values for *t-tests* (*Wilcoxon tests*)
EQSC > **EQS**	0.0974* (*0.0768**)
SPE&APE > **SPEC**	0.1351 (*0.0961**)
EQSC > **SPEC**	0.0146** (*0.0145***)
EQSC ≠ **SPEC**	0.0292** (*0.0290***)
EQS > **SPE&APE**	0.4543 (*0.5363*)
EQS ≠ **SPE&APE**	0.9086 (*0.9395*)

The results on motivation and satisfaction stress that classmates tend to respond negatively to peer evaluations. As a result, classmates are more likely to lose their motivation on the task as a consequence of peer evaluations. If we interpret friendship as a source of intrinsic motivation that would lead to an increase in satisfaction and motivation in teamwork (Lepper 1973, Lepper, Greene and Nisbett 1973) then the reduction in satisfaction that follows from the use of peer evaluations could be interpreted as evidence for the crowding-out hypothesis of intrinsic motivation among fellow subjects.

We now focus on the peer assessments released by team partners in order to assess the motives underlying the allocation of the joint outcome and then shed light on the mechanisms underlying the impact of peer evaluations on team performance.

3.5 Equity and Equality norms in peer assessments

We analyze in detail the motives that underlie the choice of the allocation rule for the team outcome in peer evaluations treatments. We stress the fact that two competing norms drive subjects' choices to allocate the joint outcome whereas very few subjects follow pure self-interest.

In order to analyze the different factors underlying subjects' peer assessments decisions we first define the concepts of claims that are consistent with pure self-interest and equality concerns.

Definition 1. *Pure self-interest. Peer evaluations respond to self-interest if in treatment (***SPE***) we observe* $\alpha_{ii} = 100\%$. *In treatment (***APE***) subjects* L *and* R *are purely self-interested if* $\alpha_{LM} = 0$ *and* $\alpha_{RM} = 0$, *respectively.*

A subject uses the equality norm when he decides to split the future joint outcome evenly independently of the relative contributions. We define this concept below.

Definition 2. *i) Strict equality. Peer evaluations are consistent with strict equality in treatment (***SPE***) if* $(\alpha_{iL}, \alpha_{iM}, \alpha_{iR}) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ *and in treatment (***APE***) if* $\alpha_{jM} = \frac{1}{3}$ *for subjects* $j \in \{L, R\}$*. For subject* M*, strict equality applies in treatment (***APE***) if* $\alpha_{ML} = \alpha_{MR} = \frac{1}{2}$*.*

*ii) Weak equality. Peer evaluations are consistent with weak equality in treatment (***SPE***) if the same share of the joint outcome is assigned to both team partners, that is* $\alpha_{ij} = \alpha_{ik}$ *for* $i \neq j \neq k$*. In treatment (***APE***) strict and weak equality coincide.*

In order to assess the relevance of the equity norm we use data on subjects' performance in Task 2 as a measure of the relative contribution of each team member in Task 1.[24] We have to keep in mind the extreme similarity between the two tasks implying that a partner's performance on Task 2 is likely to be an accurate signal for the subject's contribution to the performance obtained in the collective Task 1.

Definition 3. *Relative contribution. If subjects* L*,* M *and* R *achieve performance levels* x_L*,* x_M *and* x_R *when completing Task 2 individually, we compute subject* i*'s relative contribution to the team output as* $\rho_i = \frac{x_i}{x_L + x_M + x_R}$*.*

Definition 4. *Equity. Peer evaluations are consistent with equity when one assigns a larger share of the joint outcome to the team partner with a higher relative contribution. That is,* $\alpha_{ij} \geq \alpha_{ik}$ *whenever* $\rho_j \geq \rho_k$ *for* $i \neq j \neq k$*.*

This definition of equity follows directly from the *proportionality principle* under which fair allocations should depend on subjects' inputs determined by both effort and ability. We consider that a subject uses the equity norm when

[24] This measure of relative contribution is actually consistent with subjects' assessments of each team partner's contribution reported at the end of Stage 2 in the case of treatment (**APE**).

he decides to assign the future team outcome with respect to his knowledge of the relative contribution of the different team partners. In our experimental design subjects' relative contributions depend both on their level of effort and on their ability levels. In that respect, our definition of the equity norm is weaker than the definition of equity that follows from the *accountability principle* under which fair rewards are in proportion to the contributions that individuals control (Konow 2000, Konow 2003, Konow, Saijo and Akai 2008). According to this principle, innate abilities should not affect the fair allocation of the group outcome. We allow for a broader definition of the equity norm in line with the *proportionality principle*.[25]

We summarize our findings on the relative importance of either norm in the following result.

Result 3. (Equity and Equality norms and peer evaluations)

Even though a majority of peer assessments are driven by equality concerns, a significant proportion of assessments are also consistent with equity concerns.

Support for Result 3.

The relative importance of the equity and equality norms is displayed in the following table based on the consistency of subjects' peer evaluations with definitions 1 to 4.

[25] We argue that effort and ability tend to be highly correlated as subjects having a special talent to find numbers satisfying certain restrictions will also try to exert more effort to take advantage of their innate ability. In general, highly-talented individuals can be seen as having low marginal costs of effort. As a result, under the classical assumption of convex cost function highly-talented individuals will exert high levels of effort.

Table 14: Classification of subjects according to their peer evaluations

	Strict Egalitarian	Weak Egalitarian	Equity	Self-interest
SPE	7	7	6	4
APE	12	-	9	0
SPEC (classmates)	20	5	2	3
E-SPEC (classmates)	25	0	13	1
Total	64	12	30	8
(% of the total)	(56.1%)	(10.5%)	(26.3%)	(7.1%)

The prevalence of the equality norm follows from the fact that a majority of subjects (56.1%) chose a strictly egalitarian allocation rule giving one third of the joint output to each team member. We find strong egalitarianism concerns in the treatments involving classmates whether they self-selected into teams or not. We disregard the possibility that the high level of egalitarianism observed with classmates is due to a lower level of dispersion in subjects' abilities.[26] We find that the average dispersion in subjects' abilities for a given team is not significantly different between teams composed of strangers compared to teams composed of classmates.[27]

These results emphasize that the use of the equality norm in the allocation of team outcomes is not restricted to the focal split 50-50 (Andreoni and Bernheim 2009, Corgnet, Sutan and Veszteg 2009). Also, a significant proportion of subjects' allocation rules are consistent with equity. The extensive use of the equity norm implies that subjects have an interest to appear as hard team workers in order to obtain a large share of the team outcome. Indeed, if strict

[26] Dispersion of partners' abilities is measured as the standard deviation of team partners' performance on the individual task completed in Stage 2.

[27] We find that the p-value is equal to 0.3210 for the t-test and it is equal to 0.3873 for the Wilcoxon rank-sum test.

equality was the only relevant factor in explaining subjects' peer assessments then each subject would be rewarded with certainty one third of the team outcome as in the case of the equal split treatment. In that case as well as in the case of pure self-interest we would expect peer evaluations to have no effect on either motivation or team performance.[28]

4 Discussion

In this paper we have emphasized that using sharing rules for team outcomes that depend on peer evaluations instead of using equal splitting rules decreased team performance. In particular, we found that peer evaluations undermine team performance when team members know each other. We accounted for this result by stressing that peer evaluations are likely to weaken the high level of intrinsic motivation that is usually observed in teams formed by friends.

We finally identified the main determinants of peer evaluations emphasizing that the majority of subjects (56.1%) exhibit equality concerns and decide to split the team outcome equally among partners. However, a significant proportion of subjects (26.3%) base their peer evaluations on their partners' relative contribution. In that case, peer evaluations can possibly affect subjects' motivation and team performance since partners have an incentive to appear as high contributors.

As an agenda for future research, one could attempt to establish institutional frameworks in which the commonly used equal splitting rule is dominated by peer evaluations rules. This may be the case in a context in which free riding is particularly high as is likely to occur if tasks are complex or if

[28] One may argue that peer evaluations systematically increase peer pressure. However, if subjects are driven by strict egalitarian concerns why would they be more attentive to their partners' relative contribution in the peer evaluations treatments than in the equal split treatment.

teams are large. We may also analyze whether subjects' motivation and team performance can be improved by letting team partners choose whether to use an equal splitting rule or peer evaluations.

5 Appendix

Instructions for the Baseline treatment

(EQS)

The goal of this experiment is to study individual decision making. The instructions are simple and if you follow them carefully you will receive a considerable amount of money in cash by the end of the experiment. Payments will be made confidentially, so no one will receive information about the earnings of the other participants. You can ask a question at any time by raising your hand first. Apart from these questions it is strictly forbidden to talk among participants. You are only allowed to talk with your teammates. Talking with another subject may result in immediate expulsion from the experiment.

This experiment consists of two tasks. In all of them you will have to work alone or with your assigned partners on a common task for which you will be provided a pencil and a sheet of paper. All participants have received the same instructions as you.

For each task, you will be asked to find, in a limited amount of time, numbers that satisfy specific conditions.

Example: Find as many numbers as you can, satisfying the following conditions:

- It has 3 digits.
- If you sum its digits the result is equal to 7.
- If you multiply its digits the result is strictly larger than 10.
- The number 124 is not a solution, because although it has 3 digits and they

sum up to 7 ($1+2+4=7$), if you multiply its digits the result is 8 ($1\times 2\times 4=8$) which is smaller than 10.

- The number 423 is not a solution either, because although it has 3 digits and if you multiply them the result is larger than 10 ($4\times 3\times 2=24$), the sum of its digits is not 7, but 9 ($4+2+3=9$).

- The number 322 is a solution, because $3+2+2=7$ and $3\times 2\times 2=12$, which is higher than 10.

For each number that you find you will be rewarded with 45 euro cents and for each incorrect answer you will lose 25 euro cents. Your total profits can never be negative. As a result, if you and your partner find n numbers together, you earn n times 45 euro cents. Your individual earnings will be computed as the half of your group earnings, for this reason your payoff will be equal to n times 45/3 euro cents.

Also, each wrong answer entails a penalty of 45 euro cents. If you get much less correct answers than incorrect answers your benefits are zero.

Notice that by answering correctly you can earn a considerable amount of money in a short period of time!

Stage 1. - Task 1 -

You have 18 minutes to find as many numbers as you can, satisfying the following conditions:

- It has 3 or 4 digits.
- If you sum its digits the result is equal to 15.
- If you multiply its digits the result is strictly larger than 10.
- The last two digits are strictly larger than 1.
- The first digit is an odd number.
- The second digit is an even number.

Stage 2. (Subjects are separated and isolated)

- Questions -

Please answer now the following questions. The answers to these questions are anonymous and will never be revealed along the experiment.

1. What has been your individual contribution, in percentage terms, to the performance of the group?

2. What is the percentage of the group profits that you would claim in order to undertake a similar task with the same person?

- Task 2 -

You have 6 minutes to find as many numbers as you can, satisfying the following conditions:

- It has 3 or 4 digits.
- If you sum its digits the result is equal to 14.
- If you multiply its digits the result is strictly larger than 10.
- The last two digits are strictly larger than 1.
- The first digit is an odd number.
- The second digit is an even number.

For each number that you find you will be rewarded with 30 euro cents and for each incorrect answer you will lose 15 euro cents.

Instructions for the Symmetric Peer Evaluations Treatment (SPE)

The only difference with the baseline treatment is the payment scheme. Subjects knew before starting the team task that they would be paid according to peer evaluations as follows.

- Individual Earnings.

To compute individual earnings, each team member will decide anonymously a sharing rule for the team outcome obtained in the first task. Make sure that your numbers add up to 100%. The actual share obtained by a team member will be computed as the average of the three sharing rules chosen by the three team partners.

- Example.

Consider the case of Team 1. Subject $R1$ assigns 25% of the total earnings to $L1$, subject $M1$ assigns 50% to subject $L1$ whereas subject $L1$ assigns 45% of the total earnings to himself. Then, subject $L1$ obtains a share of the team profits equal to: $\frac{25\% + 50\% + 4\ 5\%}{3} = 40\%$.

Subject	Assigned share of the team earnings
L	
M	
R	
Total	100%

Instructions for the Asymmetric Peer Evaluations Treatment (APE)

The only difference with the baseline treatment is the payment scheme. Subjects knew before starting the team task that they would be paid according to the following rule.

- Individual Earnings.

If you are subject L or R, you decide anonymously upon the share of the team outcome assigned to subject M and subject M will decide how to share the remaining team profits between subjects L and R.

If you are subject L or R, you have to complete the following table with a number between 0 and 100%. The share of the joint profits actually obtained by subject M will be computed as the average of the shares assigned by subjects L and R to subject M.

Subject	Assigned share of the team earnings
M	

If you are subject M, you assign anonymously the remaining joint profits to subjects L and R respectively by completing the following table making sure that the sum of the two assigned shares sum up to 100%.

Subject	Assigned share of the team earnings
L	
R	
Total	100%

- Example. If subjects L and R decide to assign 20% and 40% to subject M, respectively, then subject M obtains a share of the team profits equal to $\frac{20\%+40\%}{2} = 30\%$. In addition, if subject M assigns 40% to subject L and 60% to subject R, then the share of joint profits obtained by subjects L and R are respectively equal to $(1 - 30\%) \times 40\% = 28\%$ and $(1 - 30\%) \times 60\% = 42\%$.

6 References

Alchian, A., H. Demsetz. 1972. "Production, information costs, and economic organization" *American Economic Review* 62, 777-795.

Andreoni, J., B. Bernheim. 2009. "Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects" forthcoming *Econometrica*.

Baker, G., Gibbons R., K. Murphy. 1994. "Subjective performance measures in optimal incentive contracts" *Quarterly Journal of Economics* 109, 1125-1156.

Bartlett, R. 1995. "A flip of the coin - A roll of the die: An answer to the free-rider problem in economic instruction" *Journal of Economic Education* 26, 131-139

Bénabou, R., J. Tirole. 2006. "Incentives and prosocial behavior" *American Economic Review* 96, 1652-1678.

Bochet, O., Page, T., L. Putterman. 2006. "Communication and punishment in voluntary contribution experiments" *Journal of Economic Behavior and Organization* 60, 11-26

Booth, A., J. Frank. 1999. "Earnings, productivity, and performance-related pay" *Journal of Labor Economics* 17, 447-463.

Calsamiglia, C., Franke, J., P. Rey Biel. 2009. "The incentive effects of affirmative action in a real effort tournament" *mimeo.*

Clark, N., Davies, P., R. Skeers. 2005. "Self and peer assessment in software engineering projects" Proceedings of the 7th Australasian conference on Computing education volume 42.

Colquitt, J., Zapta-Phelan, C., Q. Roberson. 2005. "Justice in teams: A review of fairness effects in collective contexts" *Research in Personnel and Human Resources Management* 24, 53-94.

Conway, R., Kember, D., Sivan, A., M. Wu. 1993. "Peer assessment of an individual's contribution to a group project" *Assessment and Evaluation in Higher Education* 18, 45-56.

Corgnet, B., Sutan, A., R. Veszteg. 2009. "Equity and equality norms in team-formation experiments: Evidence from France, Japan and Spain" *mimeo.*

Devine, D., Clayton, L., Philips, J., Dunford, B., S. Melner. 1999. "Teams in organizations: prevalence, characteristics and effectiveness" *Small Group Research* 30, 678-711.

Dickinson, D., M. Isaac. 1998. "Absolute and relative rewards for individuals in team production" *Managerial and Decision Economics* 19, 299-310.

Dickinson, D., M. Villeval. 2008. "Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories" *Games and Economic Behavior* 63, 56-76.

Dochy, F., Segers, M., D. Sluijsmans. 1999. "The use of self-, peer-, and co-assessment in higher education: A review" *Studies in Higher Education* 24, 331-350.

Dohmen, T., A. Falk. 2006. “Performance pay and multi-dimensional sorting: Productivity, preferences and gender” *IZA Discussion Paper No. 2001.*

Encinosa, W., Gaynor, M., J. Rebitzer. 2007. “The sociology of groups and the economics of incentives: theory and evidence on compensation systems” *Journal of Economic Behavior and Organization* 62, 187-214

Falk, A., A. Ichino. 2006. “Clean evidence on peer effects” *Journal of Labor Economics* 24, 39-58.

Falk, A., M. Kosfeld. 2006. “The hidden costs of control” *American Economic Review* 96, 1611-1630.

Fehr, E., S. Gächter. 2000 “Cooperation and punishment in public goods experiments” *American Economic Review* 90, 980–994.

Fehr, E., B. Rockenbach. 2003. “Detrimental effects of sanctions on human altruism” *Nature* 422, 137-140.

Frey, B. 1993. “Does monitoring increase work effort? The rivalry with trust and loyalty”. *Economic Inquiry* 31, 663-70.

Gaynor, M., M. Pauly. 1990. “Compensation and productive efficiency in partnerships: evidence from medical group practice” *Journal of Political Economy* 98, 544-573.

Gneezy, U., Niederle, M., A. Rustichini. 2003. “Performance in competitive environments: gender differences” *Quarterly Journal of Economics* 118, 1049-1074.

Güth, W., Schmittberger, R., B. Schwarze. 1982. “An experimental analysis of ultimatum bargaining” *Journal of Economic Behavior and Organization* 3, 367-388.

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

Irlenbusch, B., G. Ruchala. 2008. “Relative rewards within team-based compensation” *Labour Economics* 15, 141-167.

Isaac, M., J. Walker. 1988. “Communication and free-riding behavior: The voluntary contributions mechanism" *Economic Inquiry* 26, 585-608.

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

Konow, J. 2000. "Fair shares: Accountability and cognitive dissonance in allocation decisions" *American Economic Review* 90, 1072-92.

Konow, J. 2003. "Which is the fairest one of all? A positive analysis of justice theories" *Journal of Economic Literature* 41, 1186-1237.

Konow, J., Saijo, T., K. Akai. 2008. "Morals and mores? Experimental evidence on equity and equality from the US and Japan" *mimeo.*

Kozlowski, S., Gully, S., Nason, E., E. Smith. 1999. "Developing adaptive teams: a theory of compilation and performance across time and levels" In D.R. Ilgen & E.D. Pulakos (Eds). The changing nature of work and performance: implications for staffing, personnel actions, and development. *San Francisco, CA: Jossey-Bass*, 240–294.

Lawler, E., Mohrman, S., G. Ledford. 1995. "Creating high performance organizations: practices and results of employee involvement and total quality management in fortune 1000 companies" *San Francisco, CA: Jossey-Bass.*

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

Lazear, E., S. Rosen. 1981. "Rank-order tournaments as optimum labor contracts" *Journal of Political Economy* 89, 841-864.

Leibowitz, A., R. Tollison. 1980. "Free-riding, shirking and team production in legal partnerships" *Economic Inquiry* 18, 380-394.

Lepper, M. 1973. "Dissonance, self perception and honesty in children" *Journal of Personality and Social Psychology* 25, 65-74.

Lepper, M., Greene, D., R. Nisbett. 1973. "Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis" *Journal* *of Personality and Social Psychology* 28, 129-137.

Levin, J. 2003. "Relational incentive contracts" *American Economic Review* 93, 835-847.

MacLeod, B. 2003. "Optimal contracting with subjective evaluation" *American Economic Review* 93, 216-240.

Masclet, D., Noussair, C, Tucker, S., M. Villeval. 2003. "Monetary and nonmonetary punishment in the voluntary contributions mechanism" *American Economic Review* 93, 366-380

Milgrom, P., J. Roberts. 1992. "Economics, organization and management" *Prentice Hall.*

Montmarquette, C., Rulliere, J., Villeval, M., R. Zeiliger. 2004. "Redesigning teams and incentives in a merger: An experiment with managers and students" *Management Science* 50, 1379-1389.

Nalbantian, H., Schotter, A. 1997. "Productivity under group incentives: an experimental study" American Economic Review 87, 314-341.

Newhouse, J. 1973. "The economics of group practice" *Journal of Human Resources* 8, 37-56.

Prendergast, C. 1993. "A theory of yes men" *American Economic Review* 83, 757-70.

Prendergast, C. 1999. "The provision of incentives in firms" *Journal of Economic Literature* 37, 7-63.

Rajan, M., S. Reichelstein. 2006. "Subjective performance indicators and discretionary bonus pools" *Journal of Accounting Research* 44, 585-618.

Ramsey, P., M. Wenrich. 1999. "Peer ratings an assessment tool whose time has come" *Journal of General Internal Medicine* 14, 581–582.

Rotemberg, J. 1994. "Humans relations in the workplace" *Journal of Political Economy* 102, 684-717.

Rotemberg, J., 2006. "Altruism, reciprocity and cooperation in the workplace" The Handbook of the Economics of Giving, Altruism and Reciprocity Part III, 21, North-Holland.

Sally, D. 1995. "Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992" *Rationality and Society* 7, 58-92.

Sefton, M., Shupp, R., J. Walker. 2007. "The effect of rewards and sanctions in provision of public goods" *Economic Inquiry* 45, 671-690.

Sutter, M. 2006. "Endogenous versus exogenous allocation of prizes in teams-Theory and experimental evidence" *Labour Economics* 13, 519-549.

Thomas, P., Gebo, K., D. Hellmann. 1999. "A pilot study of peer review in residency training" *Journal of General Internal Medicine* 14, 551-554.

Tu, Y., M. Lu. 2005. "Peer-and-self assessment to reveal the ranking of each individual's contribution to a group project" *Journal of Information Systems Education* 16, 197-205.

Van Dijk, F., Sonnemans, J., F. van Winden. 2001. "Incentive systems in a real effort experiment" *European Economic Review* 45, 187-214.

Van Rosendaal, G., P. Jennett. 1992. "Resistance to peer evaluation in an internal medicine residency" *Academic Medicine* 67, 63.

Wasserman, N. 2007. "Split decisions: how assembly, allocation, and acceleration choices affect turnover in top management teams" *mimeo.*

Weber, R. 2006. "Managing growth to achieve efficient coordination in large groups" *American Economic Review* 96, 114-126.