Gender Interactions in Team Production: Performance and Punishment*

Seeun Jung[†] & Radu Vranceanu[‡]

February 24, 2015

Abstract

This paper reports results from a real-effort experiment in which men and women are paired to form a two-member team and asked to execute a real-effort task. Each member receive an equal share of the team's output. Workers who perform better than their partner can punish him/her by imposing a fine. We manipulate the teams' gender composition (man-man, man-woman, and woman-woman) to analyze whether an individual's performance and sanctioning behavior depends on his/her gender and the gender interaction within the team. The data show that men perform better than women but that women improve their performance if paired with women. Women sanction their partners less frequently than men do, but women are more frequently sanctioned than men. Finally, when sanctioned, women exhibit worse performance, whereas men tend to show improved performance.

Keywords: Gender discrimination, Real-effort task, Team production, Performance, Punishment.

JEL Classification: M52; J16; C91

1 Introduction

In general, the organization of economic activity within firms involves teams of individuals who jointly produce goods and services. In many cases, each individual's contribution to the final output cannot be observed. Hence in practice many firms will resort to team-based incentives to reward their employees (Lawler and Mohrman (2003); Boning et al. (2007)). Economists note that such compensation schemes are inefficient; if every team member relying on the efforts of his partners reduces his own effort, in equilibrium, all team members will follow the same strategy and output collapses. A substantial body of theoretical literature investigates what mechanisms can contain free-riding behavior in team production.[1] One established result in this literature is that free-riding can be alleviated if the group develops internal monitoring processes and can punish defectors

*Seeun Jung worked on this project with the support of the Labex MME-DII (http://labex-mme-dii.u-cergy.fr). The authors would like to thank the ESSEC Research Center for funding this research and Delphine Dubart from the ESSEC Experimental Lab for her technical assistance.

†ESSEC Business School and THEMA. PB 50105, 95021 Cergy, France. E-mail: jung@essec.edu

‡ESSEC Business School and THEMA. PB 50105, 95021 Cergy, France. E-mail: vranceanu@essec.fr

[1]See for instance: Alchian and Demsetz (1972), Holmstrom (1982), McAfee and McMillan (1991), Itoh (1991), Kandel and Lazear (1992), Legros and Matthews (1993), Barron et al. (1997), Che and Yoo (2001)

by taking actions that will diminish defectors' payoffs. Such sanctions need not be monetary; mockery, intimidation and social exclusion of those who deviate from the group's norm are ubiquitous phenomena in modern organizations (Fehr and Gächter (2000)). However, psychologists warn that sanctions should be applied with caution, because they generate negative emotions that might hamper performance, which has been termed the "motivation crowding-out" effect (for a survey, see Frey and Jegen (2001) and Festré and Garrouste (2014)).[2]

One important characteristic of production teams is their gender composition. Men and women might present different behavioral responses to incentives (positive or negative), and more important, their behavior may change depending on the gender with which they are paired. One can distinguish between gender-specific characteristics and gender interactions. For instance, at the individual level, Eagly and Steffen (1986) and Eckel and Grossman (2008b) note that differences in reproductive and family roles may lead to the valuing of nurturing and conciliating qualities in women and of competitive qualities in men. It is generally accepted that women are more averse to risk than men (Eckel and Grossman (2008a); Croson and Gneezy (2009)), but this result has recently been challenged based on the weakness of the statistical tests used to confirm it (Filippin and Crosetto (2014)). Another well-documented observed behavior is that women are more averse to competition than men (Niederle and Vesterlund (2007, 2011); Croson and Gneezy (2009)).[3] Ortmann and Tichy (1999) use a Prisoner's Dilemma game to show that women have a greater propensity to cooperate than men, but their desire to cooperate fades when the game is repeated.

Turning to gender interaction in team production, psychologists argue that women may exhibit greater solidarity with women than with men (Tajfel (1981)) and, conversely, that women may regard other women as competitors (Buss (1998); Campbell (1999); Kanazawa (2005)). Gneezy (2003) reports that, when the compensation scheme becomes more competitive, women increase their performance only when they compete with women, but not when they compete with men. Differences in how men (women) react when paired with same (opposite) gender partners have been observed in ultimatum games and dictator games. For instance, Ben-Ner et al. (2004) conducts several dictator game experiments in which women and men are allowed to divide ten dollars with a completely unknown person, or a person of a known gender. They found that women give systematically less to women than to men and persons of an unknown gender. Dufwenberg and Muren (2006) show that women receive more favourable treatment than men in the dictator game. Eckel and Grossman (2001) find that women make more generous offers than men in a standard ultimatum game, that offers made by women do not depend on the partner's gender, and that they are more likely to be accepted.

Whether a group's gender composition affects individual and group performance is an important research question, especially for the management of the firms that are moving from all-man to mixed gender work-teams. Yet, as Kuhn and Villeval (2013) emphasize, "compared to the literature on gender and tournaments, the economics literature on gender and teams is sparse". As notable exceptions, Apesteguia et al. (2012) analyze

[2]The potentially detrimental effect of sanctions and the stimulating effect of refraining from imposing sanctions, have already been emphasized by experimental economists who studied the principal/agent game (Dickinson and Villeval (2008); Kirstein (2008)), the trust game (Fehr and List (2004); Fehr and Rockenbach (2003); Houser et al. (2008)) or the ultimatum game (Gneezy (2003)).

[3]Niederle and Vesterlund (2007) build their analysis on two-stage real-effort task experiment (solving mazes). Although there are no gender differences in performance in the first stage, in the next stage, men select a tournament payment scheme rather than piece-wise pay twice as often as do women.

information from a large database on a well-known business game played by self-selected teams of three students (*StratX-l'Oréal*). They show that, even when controlling for personal characteristics, all-women teams perform worse than all-men and mixed teams. A similar result is obtained by Lamiraud and Vranceanu (2014) using panel data from a different business game (*Kalystée-l'Oréal*) with random allocation of students to teams of five. Hoogendoorn et al. (2013) collect data on companies created and operated for one academic year by students enrolled in the entrepreneurship program of the Amsterdam College of Applied Sciences and find that business teams with an equal gender mix perform better than male-dominated teams in terms of profits and sales. Kuhn and Villeval (2013) use a real-effort experiment to analyze gender differences regarding self-selection in teams in a cooperative work environment.[4] They show that women have a preference for team work, whereas men would prefer individual tasks. One of the many interesting results revealed by this research is that in exogenously assigned teams there is no gender difference in performance and neither men nor women free ride.

The aim of the present paper is to contribute to the literature on gender discrimination in teams by studying performance, sanctioning strategies and the behavioral responses of sanctioned workers paired with a same/different gender partner in a real-effort task. At the outset of the experiment, subjects were randomly assigned to mixed, man-woman (MW), all-men (MM) and all-women (WW) teams, in a standard between-subjects design, where a subject cannot participate in more than one experimental session. We used a smooth communication method to ensure that team members were aware of their partner's gender. Participants were asked to count the 7s in blocs of figures successively displayed on the computer screen during six successive rounds of four minutes each. This task was introduced by Mohnen et al. (2008) and Pokorny (2008) who note that it is interesting for experimental research because it does not require any particular skill or computing ability. At each round, individuals receive an equal share of the team's output, which is proxied by the total number of correct answers.[5] In our experiment, as in Mohnen et al. (2008), if a participant chooses to rest, he/she can press a button, and the screen is blocked for a 20s period for which he/she receives a small lump-sum payment. Payoffs are calibrated such that without punishment individuals have an incentive to free ride (but they do not have an incentive to cheat, because they are paid for resting, but not for providing an incorrect answer). However, in this paper, we follow Vranceanu et al. (2014) and authorize the best performer on a team to punish the less productive partner by applying a monetary sanction. Punishment is costly for the punisher (Fehr and Gächter (2000); Falk and Fischbacher (2005); Nikiforakis (2008)).

The experiment allows us to study gender differences and gender interactions in explaining individual performance, as well as the consequences that sanctions have for individual performance. Sanctioning behavior is another important topic that can be addressed through this experiment. The results will be presented in the most straightforward manner, as a list of twenty-two key findings following the various empirical analyses. In essence, the data indicate that, in this specific task, men perform (slightly) better than women but that women perform better if paired with women. When they outperform

[4]Their experiment builds on a coding task When they choose team production, subjects guess the gender of the partner by learning his/her first name. Partners on teams receive an equal split of the joint output.

[5]Notice that this compensation scheme should encourage cooperation, as in the experiment by Kuhn and Villeval (2013), and would part from the competitive environment specific to the experiments by Gneezy (2003), or Niederle and Vesterlund (2007, 2011), in which individual pay is based on a tournament.

their partner, women sanction their partners less than men do, but women are sanctioned more frequently than men. Finally, when sanctioned women's performance tends to decline whereas men's performance improves.[6] These differences in performance / sanctioning behavior cannot be explained by gender differences in risk aversion, insofar as the task of counting 7s does not involve any form of risk-taking. Men's greater preference for competition should also not play a role because the compensation scheme (equal division of total output) promotes cooperation among team members (Kuhn and Villeval (2013)).

The paper is organized as follows: The next section introduces the experimental design. Section 3 presents the results. The final section presents our conclusions and provides some managerial implications, with all of the caveats related to the challenge of extrapolating from such simple experiments conclusions deemed to be relevant for managers in "nuts and bolts" firms.

2 Experimental design

All subjects were recruited from the student population of the ESSEC Business School (France), who answered an advertisement for paid decision experiments.[7] Five sessions were organized at the ESSEC Experimental Lab with a total of 100 subjects in October 2014, November 2014 and January 2015 (Table 1). At the beginning of the experiment, subjects are matched in pairs at random. The team composition is not changed across rounds. Interaction is anonymous, and hence subjects do not know who their partners are. They play the game on a computer screen and cannot establish eye contact with one another; instructions (provided in the Appendix) and data collection are computerized, and the program was developed using z-Tree (Fischbacher (2007)).[8]

In a typical round, subjects are asked to count the number of 7s in blocks of random numbers, successively displayed on the computer screen over four minutes (Mohnen et al. (2008); Pokorny (2008); Vranceanu et al. (2014)). The typical bloc has 30 columns and 6 rows (see Appendix); in each bloc, the number of 7s varies at random between 11 and 24, with an average of 18. Note that difficulty of the task depends on the total number of figures in a bloc (180), and not on the number of 7s, and thus it should not vary from one bloc to another. Pairs must execute the real-effort task for six consecutive rounds.

Payoffs are denominated in Experimental Currency Units (ECU) that will be converted into cash at the end of the experiment. At each point in time, subjects face a choice between two options: they can choose to work on the counting task, or they can choose to press a time-out button and, for 20 seconds, the screen is blocked, but they are still paid 6 ECUs. This second option can be viewed as an opportunity cost of working.[9]

For each round, a player's reward is a linear function in the sum of the correct answers provided by the two players; if an individual presses the time-out button k times, he/she also obtains the associated payoff (k times 6 ECU). At the end of each round, players learn their own performance (number of blocks worked on, number of correct answers, number of times the time-out button was pressed), as well as the number of correct answers provided

[6] We also used the dataset of Vranceanu et al. (2014) to ensure that there is no gender effect in groups that play a similar task without knowing the gender of the partner.

[7] As "Grande Ecole" students, this group is relatively homogenous in terms of computing and intellectual abilities, age and educational background.

[8] The computer program was developed by Delphine Dubart from the ESSEC Experimental Lab.

[9] The time-out button is deactivated 20 seconds before the end of the round (this is the average time needed to count the 7s in the last block of numbers).

by their partner. Then, the player who out-performed (i.e., provided more correct answers than the other) is asked whether he/she wishes to impose a penalty on his/her partner. If he/she answers "yes", he/she can penalize the other player by an amount p, where p varies between 1 and 30 ECUs. Punishment is costly: each unit of sanction entails a cost of 0.30 ECUs for the punisher. Such a linear punishment technology, involving a constant cost per unit of punishment has been used in many other studies (Fehr and Gächter (2000); Falk and Fischbacher (2005); Nikiforakis and Normann (2008)). In the event of equal performance, no penalty can be imposed. After the punishment decision is made, players learn their ECU payoff for the round.

Let us consider a team composed of players i and j. The "score" N_{it} (respectively N_{jt}) is the number of correct answers provided by the individual j (respectively j) in round t. Without any loss of generality, let us assume that $N_{jt} > N_{it}$. In this case, individual j can punish individual i by imposing a penalty $p_{jit} \in [0; 30]$. Compensation functions for each player are:

$$\begin{cases} Z_{it} = 10\frac{N_{it}+N_{jt}}{2} + 6k_{it} - p_{jit} \\ Z_{jt} = 10\frac{N_{it}+N_{jt}}{2} + 6k_{jt} - 0.3p_{jit} \end{cases} . \tag{1}$$

The parameters of the game were selected such that, without punishment, free-riding (i.e., pressing the time-out button) is the dominant strategy. An individual who makes a normal effort to count the 7s can provide a correct answer in 20 seconds on average. If the other player does the same, both would earn 10 ECUs. However, if one presses the time-out button (blocks the screen and rests for 20 seconds) while the other works and provides a correct answer, the player who free rides obtainss 11 ECUs (and the other receives 5ECUs). Clearly, 11 ECUs *and* resting is better than 10 ECUs *and* executing the boring task. If both players free ride and presse the time-out button, they obtain 6 ECUs.

The original contribution of this paper is to analyze the impact of the teams' gender composition on performance, and on punishment behavior. To do so, we manipulate the gender composition of the groups, to create mixed teams (MW), man-man (MM) teams and woman-woman (WW) teams. It was important to ensure that team members were aware of the gender of their partner, without conveying this information in a salient way. We thus organized sessions to which only men (only women) were invited, and mixed-gender sessions with an equal number of men and women. In the mixed gender sessions, pairs were mixed. While in the single-gender sessions participants might notice that they are paired with a same-gender partner (although we cannot be certain of this), in the mixed sessions the (different) gender of the partner could not be observed directly.

Thus, to convey information on the gender of the partner in a non-directive way and without relaxing anonymity, at the beginning of the experiment, participants were asked to complete an electronic form concerning their "personal characteristics" - age, gender and level of education. Immediately thereafter, the information was communicated to the other team member as "basic information" about his/her partner. In our student population, ages and levels of education are not differentiating characteristics. The only distinctive characteristic was gender, but no emphasis was placed on this. At the end of the experiment, we asked students whether they could recall the gender of their partner, and if so, what the gender was; 2 out of 100 subjects (one man and one women) could not recall this. This is a good indication that, while students knew the gender of their partner, they did not focus on it.

We use a typical between-subjects design, with data collected for MW, MM and WW teams separately. We ensure that students who participated in one treatment could not

Table 1: Sessions and Treatments

Date	Treatment	Number of Subjects	Number of Teams
Oct 23 2014	WW	26	13
Nov 7 2014	MW	24	12
Nov 18 2014	MM	18	9
Dec 5 2014	MW	18	9
Jan 14 2014	MM	14	7

participate in another treatment.

Table 1 presents the distribution of subjects with respect to sessions and treatments: On average, the experiment took approximately 50 minutes. At the end of the experimental session, the ECU gains were converted into cash at the exchange rate 100 ECU = 2.5 Euros. Subjects earned 15.2 Euros on average.

3 Results

3.1 Data and basic statistics

Among the 100 students who participated in the experiment, 47 were female. Thus 47 individuals were paired with a female partner. Each subject performed the task for 6 rounds; the dataset thus includes 600 observations.

Table 2 reports the summary statistics, for the whole sample (first panel) and separately for the population of men and women (second and third panels).

There are three key indicator variables that capture the gender profile of individuals and teams - FP (1 if subject is a woman), FPp (1 if the partner is a woman) and their interaction term $FP \times FPp$ (1 if the subject is a woman and is paired with a woman). Detailed definitions of all variables are provided in the Appendix.

The most precise measure of performance is the *number correct answers* provided by individual i in round t, denoted by NRA_{it}. The full-sample average performance per round is approximately 9.9 correct answers, with men performing slightly better than women (10.26 vs. 9.86); the difference is nonetheless statistically significant, a result that contrasts with findings by Kuhn and Villeval (2013) or Niederle and Vesterlund (2007) who observed no gender difference in performance, using various real-effort tasks (coding and solving mazes, respectively).

Figure 1 simply presents the evolution of the number of correct answers across rounds for both men and women. In all rounds men performed better than women. We can also observe that performance improves from rounds 1 to 5 indicating the presence of some learning effect (Benndorf et al. (2014)). One would expect the latter to fade away across rounds, being offset by fatigue and boredom, which appear to have set in round 6 when performance declines for both men and women. However, because our experiment builds on a between-subjects design, the learning effect should not be a problem.

A less precise measure of performance, but one independent of the ability to count correctly, is the *number of blocks counted* by individual i in round t, whatever the score (which can be correct or not), $NBLOCK_{it}$.

From Table 2, we observe that on average participants counted 12 blocks per round,

Table 2: Summary Statistics

	Obs	Mean	Sd	Min	Max
FE	100	0.47	0.50	0	1
FEp	100	0.47	0.50	0	1
FExFEp	100	0.26	0.44	0	1
NRA	600	9.86	3.28	0	18
NBLOCK	600	12.00	4.88	1	43
NTIMEOUT	600	0.62	1.50	0	11
DIFSC	600	0.00	4.04	-11	11
SANC	275	0.19	0.40	0	1
MSANC	275	2.87	7.49	0	30
MSANC *cond.* SANC=1	53	14.91	10.62	1	30
	Male Sample				
FEp	53	0.40	0.49	0	1
NRA	318	10.26	3.43	0	18
NBLOCK	318	13.00	5.88	1	43
NTIMEOUT	318	0.64	1.72	0	11
DIFSC	318	0.60	4.29	-11	11
SANC	171	0.22	0.41	0	1
MSANC	171	3.36	8.21	0	30
MSANC *cond.* SANC=1	37	15.51	11.16	1	30
	Female Sample				
FEp	47	0.55	0.50	0	1
NRA	282	9.41***	3.04	1	17
NBLOCK	282	10.88***	3.05	1	22
NTIMEOUT	282	0.59	1.21	0	8
DIFSC	282	-0.67***	3.62	-11	10
SANC	104	0.15	0.36	0	1
MSANC	104	2.08	6.08	0	30
MSANC *cond.* SANC=1	16	13.50	9.44	2	30

* p<0.10, ** p<0.05, *** p<0.01 of T-test between genders

Figure 1: Performance Evolution over Rounds by Gender

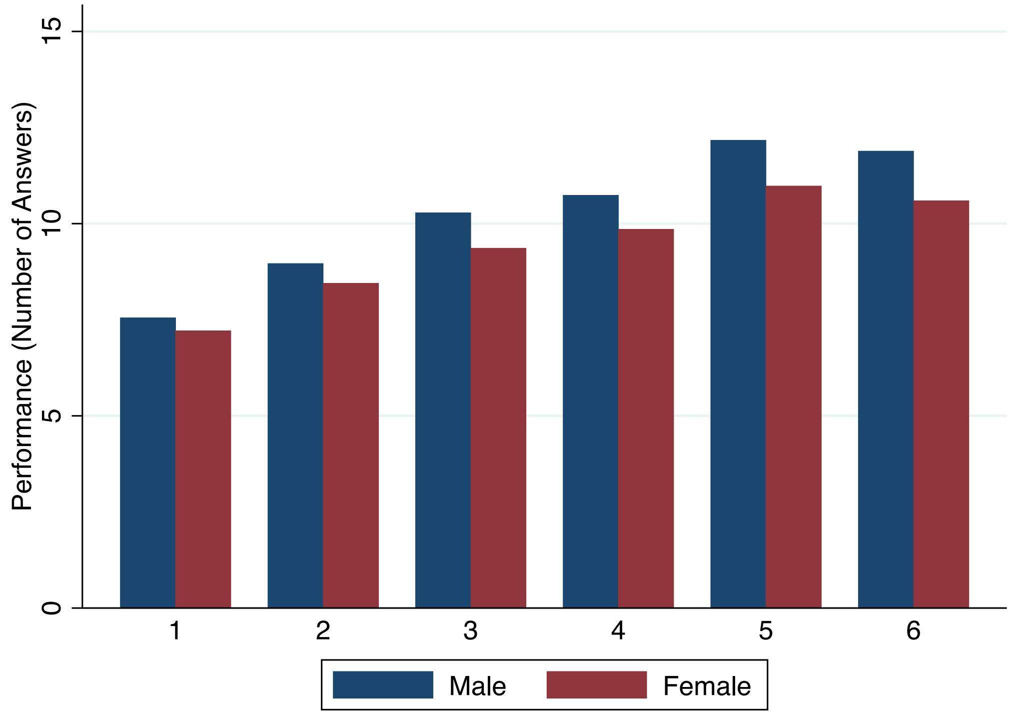


Figure 2: Trend in Sanctions across Rounds by Gender

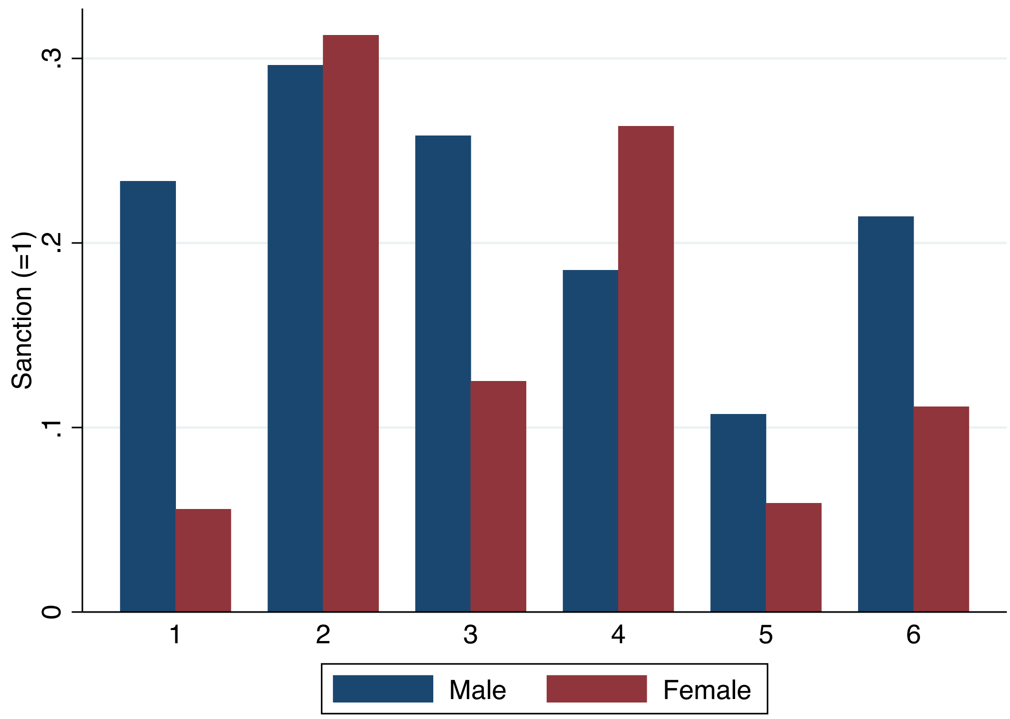
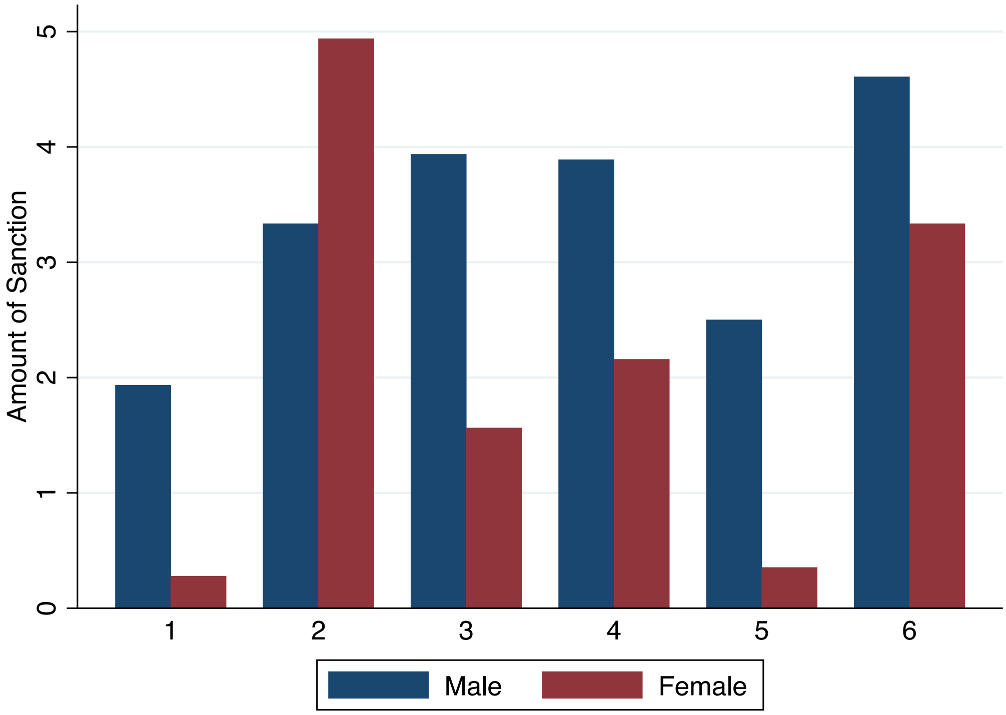


Figure 3: Change in the Sanction Amount across Rounds by Gender



with men counting 13 and women counting 10.88, the difference being statistically significant.

Using these data, it is difficult to determine whether the gender differences in performance as measured by the two former measures stem form differences in abilities (attention and speed of counting) or the deliberate choice to free-ride. Some indication is provided by the number of times that a participant chose to press the time-out button ($NTIMEOUT$), which is a good proxy for free riding behavior. With respect to this measure, the difference between men and women is statistically insignificant. Men tend to take more time-outs (0.64) than women (0.59), while their counting performance is better. This suggests that the gender differences in performance are driven by counting skills and not by alleged shirking of female subjects.

Another important measure is the *difference in the number of correct answers* between the two partners (i, j) in a round t, $DIFSC_t = NRA_{it} - NRA_{jt}$. By definition of this variable, the full-sample average difference in performance between partners is 0 (a positive performance of individual i with respect to j is offset by the negative performance of j with respect to i). Simple comparisons of the data in the second and third panels in Table 2 indicate that men perform significantly better than women.

The sanctionning behavior is captured by two measures. From the design section we know that in any round the best performer in a team can, if he/she wishes, impose a monetary sanction on his/her partner in the range [0;30] ECU, at a cost of 0.30ECU per one ECU of penalty. Sanctions could be applied in 275 cases in which one individual performed strictly better than the other. Conditional on performing better, 19% of individuals impose penalties on their partners ($SANC = 1$), for an average sanction amount ($MSANC$) equal to 2.64 ECU. Conditional on imposing a fine to the partner, the average amount of the sanction is 14.91 ECU.

Table 3: Pairwise Correlation Matrix of Variables of Interest

	FE	FEp	NRA	NBLOCK	NTIMEOUT	DIFSC	SANC	MSANC
FE	1							
FEp	0.157***	1						
NRA	-0.130**	0.0632	1					
NBLOCK	-0.218***	0.0285	0.633***	1				
NTIMEOUT	-0.0164	0.00363	-0.241***	0.0142	1			
DIFSC	-0.157***	0.157***	0.615***	0.339***	-0.167***	1		
SANC	-0.0769	0.195**	-0.107	0.0375	0.117	0.227***	1	
MSANC	-0.0830	0.184**	0.0117	0.139*	0.105	0.306***	0.786***	1

* p<0.10, ** p<0.05, *** p<0.01

Figures 2 and 3 depict the evolution of the sanction amount across rounds by gender. In the second round, individuals tend to impose the highest fines, whichgradually decrease over time. One interesting point is that individuals still impose sanctions in the last round despite it being costly. An explanation for this phenomenon was provided by Camerer and Fehr (2006) who acknowledged the existence of a large proportion of "strong reciprocators", i.e., individuals who show a combination of altruistic reward and altruistic punishment, defined as the propensity to impose sanctions on others for norm violations. Such strong reciprocators will bear the cost of imposing punishment even if they gain no individual benefit.

Simple comparisons of the data in the second and third panels of Table 2 show that men tend to apply sanctions more often than women (22% compared to 15%), and charge on average a higher amount than women (15.51 ECUs compared to 13.50 ECUs) conditional on imposing a fine to the partner.

Table 3 presents the pairwise correlation matrix of our variables of interest. The two measures of performance (NRA and $NBLOCK$) exhibit a high and significant correlation (0.63). Thus, for the sake of parsimony, in the following analyses we will consider only the number of correct answers as a measure of performance.

The number of time-outs is negatively and significantly correlated (-0.24) with the number of correct answers produced by an individual. As expected, a higher difference in scores (tantamount to lower relative performance) is associated with a higher frequency of punishment and a higher sanction amount.

While women tend to sanction their partner less frequently than do men (although the difference is insignificant) and charge a lower amount, they tend to be sanctioned more often, and be charged a higher amount (as shown by the positive and significant coefficient between FEp and $SANC$ (respectively $MSANC$)). This latter correlation corroborates a standard conclusion of the gender discrimination literature.

Several regression models will allow us to move beyond these elementary findings. Throughout our experiments, there was one male and one female participants who could not recall their partner's gender. We will exclude these two persons from our subsequent analyses.

3.2 The performance regression model

To understand the determinants of performance, we follow Mohnen et al. (2008) and Vranceanu et al. (2014) and implement our first regression model in which the dependent variable is the number of correct answers (NRA_{it}), and a key explanatory variable is the past period's sanction ($MSANC_{-1}$). However, in this new analysis, our focus is on the gender composition of the teams. We, therefore, introduce three gender indicator variables: FE (1 if subject is a woman), FEp (1 if the partner is a woman) and their interaction term ($FP \times FEp$). The number of correct answers in the previous round (NRA_{-1}) is used as a control for the individual's ability in this task, including the net learning effect. Round dummies allow us to capture the residual learning/boredom effect (not captured by the coefficient on NRA_{-1}).

Because we include one-period lagged variables, the dataset comprises only observations from round 2 to round 6. It also excludes the two individuals who mis-specified their partners.[10]

Table 4 presents the output of the panel data RE regression model. The model in column 1 does not include the interaction between being a female and receiving a sanction at the previous round, whereas this interaction is included in column 2.

Result 1.

In general, female subjects individually perform significantly worse than men, according to the negative coefficient on FE

Result 2.

Irrespective of the gender of the subject, an individual's performance is worse when teamed with a female partner, according to the negative coefficient on FEp.

Result 3.

However, if a woman is paired with a woman, their performance tends to improve (but will not offset the effects cited in R1 and R2), according to the negative coefficient on $FE \times FEp$.

Result 4.

As a corollary of R3 and R4, when men are paired with women, the team's performance deteriorates substantially.

In summary, in our data, mixed teams perform the worst, and all-male teams perform the best, with all-female teams exhibiting intermediate performance.

We can turn now to the central theme of the analysis, the strategies and consequences of punishment.

Result 5.

On average, if a subject, irrespective of the gender, was sanctioned in the previous round, s/he tends to perform better in the present round. The gain in performance is proportional to the amount of the sanction

This can be inferred from the positive coefficient of $MSANC_{-1}$, which is statistically significant in model 2.

Result 6.

However, if a woman was sanctioned at the previous round, her performance in the current round will decline.

This is shown by the negative coefficient of the interaction term $FE \times MSANC_{-1}$ in model 2, compared with the positive coefficient of $MSANC_{-1}$.

It appears that the motivation crowding-out effect mainly affects female subjects, and has little bearing on men.

As a corollary of R5 and R6, we obtain:

Result 7

Men strongly respond to sanctions by increasing performance.

Because men and women's opposite responses to sanctions tend to offset one another, this explains why the coefficient of $MSANC_{-1}$ in model 1 (for which there is no differ-

[10] There are 490 observations from five rounds for for 98 individuals.

Table 4: Performance: random effect

	(1) NRA	(2) NRA
NRA(−1)	0.749***	0.749***
	(0.04)	(0.04)
MSANC(−1)	0.015	0.094**
	(0.02)	(0.04)
FE	-0.878***	-0.742**
	(0.28)	(0.29)
FEp	-0.464*	-0.464*
	(0.27)	(0.27)
FExFEp	0.964**	0.892**
	(0.40)	(0.40)
FExMSANC(−1)		-0.105**
		(0.05)
Round3	0.038	0.022
	(0.31)	(0.31)
Round4	-0.366	-0.369
	(0.32)	(0.32)
Round5	0.637**	0.666**
	(0.32)	(0.32)
Round6	-0.644*	-0.618*
	(0.34)	(0.34)
_cons	3.602***	3.552***
	(0.39)	(0.39)
chi2	639.022	649.496
N	490	490
r2_w	0.117	0.120
r2_b	0.941	0.943
r2_o	0.571	0.576

* p<0.10, ** p<0.05, *** p<0.01

Robust S.E Clustered at the individual level.

Individuals who mis-specify their partners' gender are excluded

We consider only Round2-Round6 in order to take into account the value at $t-1$

ence regarding the gender of the person who is punished) is small and not statistically significant.

Result 8.

Performance improves across rounds, but conditionally on individual abilities, the round effects are rather mixed.

Finally, we also assessed whether an excessive sanction might entail a motivation crowding-out effect by including a quadratic term of the $MSANC$ variable in the above mentioned regression models. Its coefficient is negative, which would suggest support for the motivation crowding-out theory, but it is not statistically significant.

3.3 Determinants of the free-riding behavior

As we have emphasized, the number of times that subjects pressed the time-out button provides a indication of their propensity to free ride (and take the risk of being sanctioned). On average men press the button more often than women. As $NTIMEOUT$ and NRA are negatively correlated, one could imagine that the motivational factors determining free riding can be explained similarly to those driving performance. We therefore estimate the same regression model as in the previous section, but here use $NTIMEOUT$ as the dependent variable. The sample includes the same 490 observations. Table 5 reports results of the RE panel data regression.

We can note that:

Result 9.

Better-performing individuals tend to free-ride (press the time-out button) less often in general.

This is indicated by the negative and statistically significant coefficient of NRA_{-1}.

Result 10.

Controlling for individual ability, on average subjects tend to free-ride (press the time-out button) more often when paired with a woman.

This is indicated by the positive and statistically significant coefficient of FEp.

Result 11

However, when a woman is paired with a women, she will rest less often as indicated by the negative coefficient on $FE \times FEp$, which is larger in absolute value than the coefficient on FEp.

The corollary of Result 9 and Result 10 is:

Result 12

Men paired with women free-ride (press the time-out button) most often.

Teaming a man with a women will encourage him to free-ride.

Result 13

Over time, boredom or fatigue prompts participants to press the time-out button more often.

All coefficients for rounds 4 to 6 are positive and statistically significant.

3.4 The sanction regression model

We can turn now our attention to the determinants of sanctioning behavior. To do so, we estimate several regression models, using as a dependent variable either the indicator variable $SANC$ (1 if a sanction is imposed) or the ECU amount of the sanction, $MSANC$. As dependent variables we use the same gender indicator variables as in the performance

Table 5: Determinant of Free-Riding (The Number of Times the Time-Out Button was Pressed)

	(1) NTIMEOUT	(2) NTIMEOUT	(3) NTIMEOUT	(4) NTIMEOUT
NRA(−1)	-0.094***	-0.094***	-0.097***	-0.097***
	(0.03)	(0.03)	(0.03)	(0.03)
MSANC (-1)	-0.013	-0.040	-0.011	-0.041
	(0.01)	(0.03)	(0.01)	(0.03)
FE	0.210	0.153	0.224	0.167
	(0.20)	(0.20)	(0.20)	(0.20)
FEp	0.531***	0.535***	0.528***	0.532***
	(0.19)	(0.19)	(0.19)	(0.19)
FExFEp	-0.909***	-0.883***	-0.920***	-0.894***
	(0.28)	(0.28)	(0.28)	(0.28)
FExMSANC(−1)		0.037		0.039
		(0.03)		(0.03)
FExDIFSC(−1)			0.016	0.019
			(0.03)	(0.03)
Round3	-0.039	-0.034	-0.035	-0.029
	(0.21)	(0.21)	(0.21)	(0.21)
Round4	0.480**	0.481**	0.486**	0.488**
	(0.22)	(0.22)	(0.22)	(0.22)
Round5	0.407*	0.396*	0.416*	0.407*
	(0.22)	(0.22)	(0.22)	(0.22)
Round6	0.521**	0.512**	0.535**	0.527**
	(0.23)	(0.23)	(0.24)	(0.24)
_cons	1.175***	1.196***	1.197***	1.223***
	(0.28)	(0.28)	(0.29)	(0.29)
chi2	36.664	38.201	36.903	38.557
N	490	490	490	490
r2_w	0.078	0.082	0.078	0.083
r2_b	0.006	0.001	0.004	0.000
r2_o	0.065	0.066	0.066	0.068

* p<0.10, ** p<0.05, *** p<0.01

Robust S.E Clustered at the individual level.

Individuals who mis-specify their partners' gender are excluded

We consider only Round2-Round6 in order to take into account the value at $t-1$

equation (FE, FEp and their interaction), to which we add the difference in scores between the two partners ($DIFSC$) and the interaction term $FEp \times DIFSC$. There were initially 275 observations for which the sanction option was available (i.e., the subject performed better than his/her partner); however, 5 observations from the two individuals who mis-specified their partners must be excluded. As there are 6 rounds, we can treat this dataset as panel data, thereby allowing us to use a random-effects model to control for both individuals and time effects.

Table 6 reports the estimation output. We use three different specifications for both $SANC$ and $MSANC$; columns (1) and (4) present the regression model without interaction terms, columns (2) and (5) include the $FE \times FEp$ interaction term, and columns (3) and (6) include the $FEp \times DIFSC$ interaction term.

This brings us to

Result 14.

The frequency of imposing sanctions and the amount of the sanction is positively related to the difference in the number of correct answers (performance). In other words, the worse the relative performance of the poor performer on the team is, the higher will be his/her sanction (or likelihood of receiving a sanction).

This is shown by the positive and significant coefficient on $DISFSC$ in models (1), (2), (4), and (5).

Result 15

The sanction response to poor performance is stronger when the poor performer on the team is a woman

Indeed, the coefficients of the interaction term between FEp (the partner is a woman) and $DIFSC$ (the difference in performance between team members) is significantly positive.

Result 16

When women perform better than their partner, they tend punish him/her less than men do.

This is indicated by the negative coefficient of FE (albeit this result is not significant).

Result 17

However, when women perform poorly, they tend to be penalized significantly more than men.

This is reflected by the positive coefficient of FEp (strongly significant in models (1), (2), (4), and (5)).

Result 18

As a rough result (not statistically significant), women tend to be favored by other women when the latter can impose a sanction.

This is shown by the negative coefficient of $FE \times FEp$. This might suggest that Result 9 is essentially driven by men sanctioning women who perform poorly to a greater extent than they do men who perform poorly.

Result 19

The frequency and magnitude of sanctioning is the highest in the second round.

The second round coefficients are positive and statistically significant in all models. After the second round, round effects appear to disappear.

Table 6: Sanction Determinants

	(1) SANC	(2) SANC	(3) SANC	(4) MSANC	(5) MSANC	(6) MSANC
DIFSC	0.041***	0.037***	0.018	0.876***	0.865***	0.270
	(0.01)	(0.01)	(0.01)	(0.18)	(0.18)	(0.22)
FE	-0.056	-0.034	-0.049	-1.148	-0.256	-0.877
	(0.05)	(0.08)	(0.05)	(0.91)	(1.42)	(0.88)
FEp	0.210***	0.222***	0.042	3.894***	4.386***	-1.022
	(0.05)	(0.06)	(0.08)	(0.91)	(1.11)	(1.42)
FExFEp		-0.067			-1.652	
		(0.10)			(1.81)	
FEpxDIFSC			0.049***			1.416***
			(0.02)			(0.33)
Round2	0.147**	0.140*	0.140*	2.990**	2.909**	2.765**
	(0.07)	(0.08)	(0.07)	(1.36)	(1.39)	(1.36)
Round3	0.048	0.039	0.054	1.665	1.625	1.865
	(0.07)	(0.08)	(0.07)	(1.34)	(1.37)	(1.34)
Round4	0.082	0.062	0.061	1.943	1.890	1.447
	(0.07)	(0.08)	(0.07)	(1.37)	(1.39)	(1.36)
Round5	-0.046	-0.055	-0.053	0.868	0.844	0.731
	(0.07)	(0.08)	(0.07)	(1.36)	(1.39)	(1.36)
Round6	-0.011	-0.025	-0.019	2.255*	2.145	2.033
	(0.07)	(0.08)	(0.07)	(1.37)	(1.39)	(1.36)
_cons	-0.074	-0.047	0.010	-3.397**	-3.484**	-1.131
	(0.07)	(0.07)	(0.08)	(1.39)	(1.37)	(1.39)
chi2	45.906	42.760	52.899	51.113	51.989	72.504
N	270	270	270	270	270	270
r2_w	0.158	0.148	0.169	0.157	0.155	0.189
r2_b	0.171	0.216	0.272	0.255	0.294	0.437
r2_o	0.131	0.136	0.157	0.156	0.164	0.220

* p<0.10, ** p<0.05, *** p<0.01, Robust S.E Clustered at the individual level.

5 observations from Individuals who mis-specify their partners' gender are excluded

Table 7: Preference Questionnaire about Partner/Gender

	(1) Performance Partner	(2) Performance Other Gender	(3) Punishment Other Gender
FE	-0.227	0.042	-0.233
	(0.23)	(0.21)	(0.30)
FEp	-0.283	0.220	0.440**
	(0.26)	(0.19)	(0.20)
FExFEp	0.488	-0.261	0.493
	(0.34)	(0.25)	(0.33)
FEpxAvDIFSC	-0.109	-0.038	-0.139***
	(0.07)	(0.05)	(0.05)
AvDIFSC	-0.181***	0.059	0.056*
	(0.05)	(0.04)	(0.03)
_cons	3.906***	3.000***	3.031***
	(0.14)	(0.15)	(0.15)
N	98	98	98
r2	0.474	0.083	0.245

* p<0.10, ** p<0.05, *** p<0.01
Bootstrapping 1000 replications
2 Individuals who mis-specify their partners' gender are excluded.
avDIFSC is the average of DIFSC for each individual.

3.5 The perception regression model.

As noted above, at the end of the experimental session participants were asked to answer three questions regarding their gender preference and use a five-item scale to rank their answers (from very good to very poor). The sample includes the 98 participants who knew the gender of their partner. Table 7 presents the regression models using these expressed preferences as a dependent variable. The $AvDIFSC$ is the all-round average difference in an individual's score (positive if he/she outperformed, negative if he/she under-performed), and $FEp \times AvDIFSC$ is the interaction term with the partner being female.

The first column presents the results for the question "How do you evaluate your partner's performance?". While many factors are insignificant, we find that:

Result 20

The better the relative performance of an individual, the lower the assessment of his/her partner's performance (and vice-versa).

In line with intuition, this result is confirmed by the negative coefficient of $AvDIFSC$.

Result 21

However, all things equal, subjects tend to be more critical of women than of men, by under-estimating the ex-post performance of women.

This is reflected by the statistically significant negative coefficient of the interaction term $FEp \times AvDIFSC$.

The second column pertains to the question "In this experiment, considering the performance of your partner, you believe that a [opposite gender as partner] would have performed" for which we did not observe any significant results. Based solely on the sign of coefficient of FEp, if the subject was paired with a woman, s/he tended to believe that if s/he had been paired with a man, the male partner would have performed better (and this belief is stronger for men, as shown by the sign of $FE \times FEp$). This suggests that subjects naturally underestimate the performance of women.

The final column reports the results for the question "If your partner had been of the opposite gender, how do you imagine s/he would have penalized you?".

Result 22

As shown by the positive and significant coefficient of FEp, *if the individual had been paired with a female, s/he believes that if the partner had been a man, he would have imposed heavier sanctions.*

Subjects naturally tended to believe that women would penalize less than men, which is in line with the observed behavior in this experiment.

4 Conclusion

Gender discrimination in the workplace is an important research subject, which is motivated by the worrisome situation of women in many enterprises and work environments. The aim of this paper is to contribute to this literature by developing a real-effort experiment, in which pairs of individuals were asked to "produce" correct answers in a counting task. Partners on a team receive receive an equal share of the team's output as compensation. Furthermore, the best performers on a team are given the option to impose a monetary sanction on the less-effective partner. As an original contribution, in this paper we manipulate the gender composition of the teams, and observe individual behavior in all-men, all-women and mixed teams.

The analysis provided several results, that are not independent of the specific features of the effort task and compensation scheme. In the experiment, subjects must count 7s in blocs of figures displayed successively on the computer screen (Mohnen et al. (2008), Pokorny (2008), Vranceanu et al. (2014)). Compared to other real-effort tasks used in team production experiments (such as coding and, solving mazes), this task does not require any particular reasoning skills, but draws heavily on attention, focus and resistance to boredom. , The equal split compensation scheme supports cooperation between team members. As Kuhn and Villeval (2013) note, this work-environment is preferred by women subjects, in contrast to men, who tend to prefer more competitive reward schemes.

Among the main results, we emphasize the following:

(1) Men perform better than women, but women will improve their performance if paired with women.

(2) Women sanction their partners less than men do, but women are sanctioned more often than men.

(3) If sanctioned, women's performance tends to deteriorate, whereas men tend to improve their performance.

These results can shed some light on the dynamics of performance in teams. We observed that men tended to improve their performance from one round to another when they were sanctioned, and the gain in performance is positively related to the amount of the sanction. As the sanction is proportional to the difference in scores, in the long run, the gap in the performance of the two team members should narrow to the lowest natural difference in abilities, with sanctions gradually declining and stabilizing. This corresponds to a highly efficient system of work organization. However, we have shown that women respond to sanctions by reducing their performance. Furthermore, should they under-perform, women receive larger sanctions than men, and these sanctions are the most substantial when women are paired with men. Such a system can only lead to an on-going deterioration of women's performance. Mixed teams in which the male partner has an advantage in executing the task turn are an inefficient organizational form in the context of our experiment.

There is an important managerial implication from these empirical findings. In many enterprises and organizations, production was historically primarily completed by men. This is true for many mass-production processes in manufacturing and construction, but also in many clerical jobs (consulting, law firms, and even in academia). It is easy to understand why a norm of punishing defectors should emerge in men dominated work environments. With the accelerated feminization of many organizations, the gender composition of teams is changing rapidly. Yet, if norms do not change at the same pace, and workers uniformly impose internal punishment, observations of deteriorating firm performance would not be surprising. An efficient organization of work in such teams would involve replacing punishment with forgiveness, especially when women miss the production targets.

References

Alchian, A. A. and H. Demsetz, "Production, Information Costs, and Economic Organization," *American Economic Review* 62 (December 1972), 777–95.

Apesteguia, J., G. Azmat and N. Iriberri, "The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field," *Management Science* 58 (January 2012), 78–93.

Barron, J. M., Gjerde and K. Paulson, "Peer Pressure in an Agency Relationship," *Journal of Labor Economics* 15 (April 1997), 234–54.

Ben-Ner, A., F. Kong and L. Putterman, "Share and Share Alike? Gender-pairing, Personality, and Cognitive Ability as Determinants of Giving," *Journal of Economic Psychology* 25 (October 2004), 581–589.

Benndorf, V., H. A. Rau and C. Sölch, "Minimizing Learning Behavior in Experiments with Repeated Real-Effort Tasks," *Available at SSRN: http://ssrn.com/abstract=2503029 or http://dx.doi.org/10.2139/ssrn.2503029* (2014).

Boning, B., C. Ichniowski and K. Shaw, "Opportunity Counts: Teams and the Effectiveness of Production Incentives," *Journal of Labor Economics* 25(4) (2007), 613–50.

Buss, D., *Evolutionary Psychology: The New Science of the Mind* (Boston, Allyn and Bacon., 1998).

Camerer, C. F. and E. Fehr, "When does "Economic Man" Dominate Social Behavior?," *Science* 311(5757) (2006), 47–52.

Campbell, A., "Staying Alive: Evolution, Culture, and Women's Intrasexual Aggression," *Behavioral and Brain Sciences* 22 (1999), 203–252.

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

CROSON, R. AND U. GNEEZY, "Gender Differences in Preferences," *Journal of Economic Literature* 47(2) (2009), 1–27.

DICKINSON, D. AND M.-C. VILLEVAL, "Does Monitoring Decrease Work Effort? The Complementarity Between Agency and Crowding-Out Theories," *Games and Economic Behavrior* 63 (2008), 56–76.

DUFWENBERG, M. AND A. MUREN, "Generosity, Anonymity, Gender," *Journal of Economic Behavior and Organization* 61(1) (2006), 42–49.

EAGLY, A. AND V. STEFFEN, "Gender and Aggressive Behavior: a Meta-Analytic Review of the Social Psychological Litteraturet.," *Psychol Bull* 100(3) (1986), 309–30.

ECKEL, C. C. AND P. J. GROSSMAN, "Chivalry and Solidarity in Ultimatum Games," *Economic Inquiry* 39 (2001), 171–188.

———, "Differences in the Economic Decisions of Men and Women: Experimental Evidence," *Handbook of Experimental Economics Results* 1(57) (2008a), 509–519.

———, "Men, Women and Risk Aversion: Experimental Evidence," *Handbook of Experimental Economics Results* 1 (2008b), 1061–1073.

FALK, E. F., ARMIN AND U. FISCHBACHER, "Driving Forces behind Informal Sanctions," *Econometrica* 73(6) (2005), 2017–2030.

FEHR, E. AND S. GÄCHTER, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review* 90(4) (2000), 980–994.

FEHR, E. AND J. LIST, "The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs," *Journal of the European Econonmic Association* 2 (2004), 743–771.

FEHR, E. AND B. ROCKENBACH, "Detrimental Effects of Sanctions on Human Altruism," *Nature* 422 (2003), 137–140.

FESTRÉ, A. AND P. GARROUSTE, "Theory and Evidence in Psychology and Economics about Motivation Crowding Out: A Possible Convergence?," *Journal of Economic Surveys, On-line in-print, http://dx.doi.org/10.1111/joes.12059* (2014).

FILIPPIN, A. AND P. CROSETTO, "A Reconsideration of Gender Differences in Risk Attitudes," IZA Discussion Papers 8184, Institute for the Study of Labor (IZA), May 2014.

FISCHBACHER, U., "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10 (June 2007), 171–178.

FREY, B. S. AND R. JEGEN, "Motivation Crowding Theory," *Journal of Economic Surveys* 15(5) (2001), 589–611.

GNEEZY, U., "The W Effect of Incentives," Levine's Bibliography 666156000000000315, UCLA Department of Economics, October 2003.

HOLMSTROM, B., "Moral Hazard in Teams," *Bell Journal of Economics* 13 (Autumn 1982), 324–340.

Hoogendoorn, S., H. Oosterbeek and M. V. Praag, "The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment," *Management Science* 59(7) (2013), 1514–1528.

Houser, D., E. Xiao, K. McCabe and V. Smith, "When Punishment Fails: Research on Sanctions, Intentions and Non-cooperation," *Games and Economic Behavior* 62 (March 2008), 509–532.

Itoh, H., "Incentives to Help in Multi-agent Situations," *Econometrica* 59 (May 1991), 611–36.

Kanazawa, S., "Is "Discrimination" Necessary to Explain the Sex Gap in Earnings?," *Journal of Economic Psychology* 26(2) (2005), 269–287.

Kandel, E. and E. P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy* 100(4) (1992), 801–817.

Kirstein, A., "Bonus and Malus in Principal Agent Relations with Fixed Pay and Real Effort," *Fixed Wage* 60 (2008), 280–303.

Kuhn, P. and M.-C. Villeval, "Are Women More Attracted to Co-operation Than Men?," *Economic Journal* 125 (2013), 115–140.

Lamiraud, K. and R. Vranceanu, "Group Gender Composition, Tolerance to Risk and Economic Performance: New Evidence from an Original Business Game," *mimeo, ESSEC.* (2014).

Lawler, E. E. I. and S. A. Mohrman, "Pay Practices in Fortune 1000 Corporations," *Center for Effective Organizations, publication G* 448 (2003), 3–20.

Legros, P. and S. A. Matthews, "Efficient and Nearly-Efficient Partnerships," *Review of Economic Studies* 60 (July 1993), 599–611.

McAfee, R. P. and J. McMillan, "Optimal Contracts for Teams," *International Economic Review* 32 (August 1991), 561–77.

Mohnen, A., K. Pokorny and D. Sliwka, "Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence," *Journal of Labor Economics* 26 (October 2008), 693–720.

Niederle, M. and L. Vesterlund, "Do Women Shy Away from Competition? Do Men Compete Too Much?," *Quarterly Journal of Economics* 122 (08 2007), 1067–1101.

———, "Gender and Competition," *Annual Review in Economics* 3 (2011), 601–630.

Nikiforakis, N., "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves," *Journal of Public Economics* 92 (February 2008), 91–112.

Nikiforakis, N. and H.-T. Normann, "A Comparative Statics Analysis of Punishment in Public-good Experiments," *Experimental Economics* 11 (December 2008), 358–369.

Ortmann, A. and L. K. Tichy, "Gender Differences in the Laboratory: Evidence from Prisoner's Dilemma Games," *Journal of Economic Behavior and Organization* 39 (July 1999), 327–339.

Pokorny, K., "Pay–But Do Not Pay Too Much: An Experimental Study on the Impact of Incentives," *Journal of Economic Behavior and Organization* 66 (May 2008), 251–264.

Tajfel, H., *Human Groups and Social Categories* (Cambridge University Press, 1981).

Vranceanu, R., F. E. Ouardighi and D. Dubart, "Team Production with Punishment Option: Insights from a Real-Effort Experiment," *Managerial and Decision Economics* Online at: http://dx.doi.org/10.1002/mde.2678. (2014).

APPENDIX

DESCRIPTION OF VARIABLES

FE Dummy variable: [1] if the subject is female.

FEp Dummy variable: [1] if the subject is paired with a woman.

FExFEp Dummy variable: [1] if the subject is female *and* is paired with a woman.

NRA The number of correct answers provided by the individual.

NRA(−1) The number of correct answers provided by the individual in the previous round.

NBLOCK Number of blocks counted by the subject during the round

NTIMEOUT Number of times that the time-out button was pressed by the subject during the round

DIFSC The difference between the number of correct responses of subject i and that of his/her partner j: $NRA_i - NRA_j$ in the current round

AvDIFSC The average of $DIFSC$ over the 6 rounds for an individual

SANC Dummy variable: [1] if the subject penalizes his/her partner conditional on the subject performing better than his/her partner.

MSANC The amount of penalty that the subject imposes on his/her partner when the former performs better than his/her partner.

MSANC(−1) The amount of sanction imposed by the partner at the previous round when an individual performed worse than the partner.

FExMSANC(−1) The interaction term between $MSANC_{-1}$ and FE.

FExDIFSC(−1) The interaction term between $DIFSC_{-1}$ and FE.

FEpxDIFSC The interaction between $DIFSC$ and FEp.

FEpxAvDIFSC The interaction term between FEp and $AvDIFSC$.

Round# Dummies for each round.

INSTRUCTIONS[11]

Slide 1.
Good morning. Thank you for participating in this experiment. Please read these instructions carefully and, should you have any questions, raise your hand and call the administrator. Communication between participants is forbidden. Please turn off cellular phones. A payment in cash will be provided at the end of the session.

Slide 2. Personal characteristics
-You are: [A man / A woman]
-Your age is [X]
-Your education level is: [Baccalaureate+1,+2,+3,+4+5,+6 or more years of education]

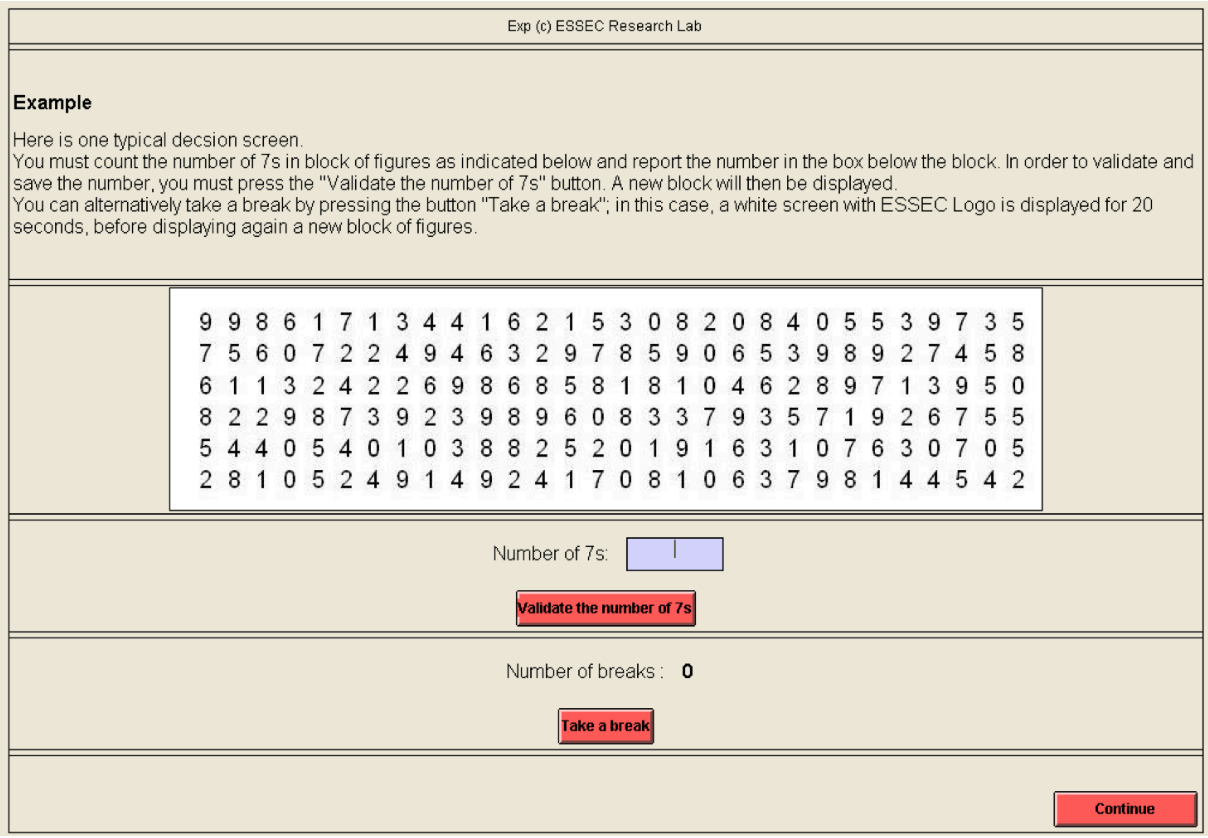
Slide 3. About your partner
In this experiment you will be paired at random with another person present in this room; this pairing is strictly anonymous.
-The gender of the partner is: [man/women]
-The age of the partner is: [X]
-The education level of the partner is: [Baccalaureate+1,+2,+3,+4+5,+6 or more years of education]

Slide 4. Main rule
-You will be required to perform an effort task jointly with a partner, during 4 identical rounds of the same experiment.
-Each round lasts for 4 minutes; the clock starts when you open the first active screen, and stops after 4 minutes. During a round, the remaining time is displayed in red characters, in the upper right corner of the screen (in seconds).
-A payoff in euros will be delivered at the end of the experiment. The payment is connected to performance in the task, according to a rule known to everyone.
-Partners will be matched in pairs at random. Your partner will not change from one round to another. His identity will not be revealed to you.
-During each round the computer displays a sequence of blocs of figures (0 to 9) in six lines and 30 columns. Your task is to accurately count how many times the figure 7 appears in a bloc and then report this number in a box. The answer is considered correct if it corresponds to the right number of 7s in the bloc, with a tolerated error margin of 1. For instance, if the correct number of 7s is 30, answers 29, 30 and 31 will be considered correct.
-Once the counted number of 7s is recorded in the box, you must press the "validate" button, to save it. After you click, a new random block of figures is automatically generated and the effort task can continue.
-At any moment you can take a break by pressing the button "Take a break". The break stops the counting task for 20 seconds; a screen with ESSEC logo appears. If the round stops in less than 20 seconds, breaks are no longer possible.
-At the end of each round the computer will display the total number of correct answers that you have provided and the total number of correct answers provided by your partner.
-At the end of each round, before moving to the next round, the player who provided the highest number of correct answers can, if he/she wants so, impose a fine on his/her part-

[11] Translated from French.

ner. The decision belongs to him/her, it is not compulsory to impose the fine. In the event that players have provided an identical number of correct answers, no sanction is possible.

Slide 5. The example slide - main decision screen



Slide 6. Compensation rule

Gains are denominated in Experimental Currency Units (ECU). For each round, the payoff for one player is made up of three elements:

1. The compensation related to the effort task
2. A gain provided when taking a break
3. Less the penalty (if any)

Which are:

1. For each player, the ECU compensation related to the effort task is equal to half the total number of correct answers provided by the team during the round, times 10.
For instance, if player 1 provided 8 correct answers and player 2 provided 5 correct answers, the gain for each player related to the effort task is 0.5x(8+5)x10=65 ECUs
2. For each 20 second break, you will receive 6 ECU, whatever your compensation for the effort task.
3. At the end of each round, before the next round starts, the player who provided the largest number of correct answers can, if he wants to impose a fine on his partner, for an amount between 1 and 30 ECUs. The gain of the partner is reduced by that amount. One ECU in fines will cost the punisher 0.30 ECU. (No sanction is possible if players provide the same number of correct answers)

-At the end of the experiment, the total amount in ECU will be converted into euros at the exchange rate 100 ECU = 2.5 euros.

Slide 8. Check questions.
To make sure that you have understood well the rules of the game, please answer these questions:

Case 1.
During the round you got 4 right answers and your partner got 2 right answers. You took two breaks.
Your gain in ECU is:
(a) 0.5*(4+2)*10+2*6; (b) (4+2)*10 + 2* 6; (c) I do not know

Case 2.
At the end of the round you got 4 right answers and your partner got 2 right answers.
Can you impose a fine on your partner?
(a) Yes (b) No (c) Don't know
Can your partner impose a fine on you?
(a) Yes; (b) No; (c) Don't know

Case 3.
At the end of the round you chose to impose a fine on your partner. The amount of the fine can be:
(a) Between 1 and 10; (b) Between 1 and 30; (c) I do not know.

Slide 9.
Correct answers are:

Case 1.
-During the round you got 4 right answers and your partner got 2 right answers. You took two breaks.
-The right answer is: you have half of the total points times 10, and the compensation for the breaks (2 x 6 ECUS), that is a total of 0.5(4+2)*10 + 2x6

Case 2.
-You got 4 right answers and your partner got 2 right answers.
-Yes, you can impose a fine on him. Attention, this is an option; you do not need to impose a fine.
-No, he cannot impose a fine on you

Case 3.
At the end of the round you chose to impose a fine on your partner. The amount of the fine can be between 1 and 30. The payoff of your partner will be reduced by this amount.

Slide 10.
-If you have any questions, please raise your hand and address it to the administrator.
-If you are sure you have understood the rules of the game, you can press the button below to launch the experiment.

-The experiment starts when all subjects have pressed the button.

Slide 11. (Main decision screen)
Similar to "Example" in Slide 5 (but without the text on top of the screen; and the timer on the right upper corner).

Slide 12. Results on task
-Your performance: Number of counted blocks [ ], Number of right answers [NRA1], Number of breaks [ ]
-The performance of your partner: Number of right answers [NRA2].
-[If NRA1 > NRA2 the computer displays] Do you want to impose a fine on the partner?
-You choose: Yes // No
-If you click Yes the computer displays "choose the amount of the fine" [A=1 to 30], then "Validate"

Slide 12. Payoff for the round
-Your partner has imposed a fine on you [or /did not impose a fine on you]
-The amount of the fine is: [A]
-Your payoff for the round is: [ ]

At the end of the experiment (after the 6 rounds)
Question 1.
Please evaluate the performance of your partner:
-[very poor, poor, average, good, very good]

Question 2.
Your partner was:
-[A man, A woman, Don't know]

Question 3.
In this experiment, considering the performance of you partner, you believe that a [opposite gender partner] would have performed:
-[much worse, worse, the same, better, much better]

Question 4.
In this experiment, do you think that a [opposite gender partner] would have applied a sanction:
-[much lower, lower, identical, higher, much higher]

Last slide
Thank you for having participated in this experiment.
The total gain for the experiment is [ ] euros.