2010

Team Formation and Self-serving Biases

Brice Corgnet
Chapman University, corgnet@chapman.edu

Recommended Citation
Corgnet, Brice. "Team Formation and Self-serving Biases." *Journal of Economics & Management Strategy* 19.1 (2010): 117-135. DOI:10.1111/j.1530-9134.2009.00247.x

Team Formation and Self-serving Biases

Comments
This is the accepted version of the following article:

Corgnet, Brice. "Team Formation and Self-serving Biases." *Journal of Economics & Management Strategy* 19.1 (2010): 117-135.

which has been published in final form at DOI: 10.1111/j.1530-9134.2009.00247.x.

Team formation and self-serving biases

October 2006

Author: Brice Corgnet
email: bcorgnet@unav.es

Universidad de Navarra
Economics Department
Edificio Bibliotecas (Entrada Este)
31080 Pamplona (Navarra), Spain

Team formation and self-serving biases

Abstract

There exists extensive evidence that people learn positively about themselves. We build on this finding to develop a model of team formation in the workplace. We show that learning positively about oneself systematically undermines the formation of teams. Agents becoming overconfident tend to ask for an excessive share of the group outcome. Positive learning generates divergence in workers' beliefs and hampers efficient team formation. This result is shown to be robust to high degrees of workers' sophistication. We finally apply our model to coauthorship and organizational issues.

1 Introduction

In recent years, more intensive use of teamwork in organizations has raised interest in the factors affecting the success of teams. In this paper, we consider the role of psychological factors on team formation and in particular we analyze the impact of self-serving biases in individuals' learning processes. The focus on agents' biased self-attribution is motivated by widespread evidence in Psychology literature showing that people tend to take credit for successes but deny responsibility for failures (Bradley 1978, Miller and Ross 1975, Zuckerman 1979). Individuals are inclined to process information distortedly so as to build a positive self-image (Fiske and Taylor 1991, Nisbett and Ross 1980). As noted by Gilbert et al. (1998):

> "Psychologists from Freud to Festinger have described the artful methods by which the human mind ignores, augments, transforms and rearranges information in its unending battle against the affective consequences of negative events."

In addition, there is extensive evidence that people recall their successes better than their failures (Korner 1950, Silverman 1964, Mischel, Ebbesen and Zeiss 1976). This leads individuals to hold excessively positive beliefs about themselves (Greenwald 1980, Svenson 1981 and Cooper, Woo and Dunkelberg 1988). Psychology literature has mostly interpreted biases in inference and attribution as motivational biases. Agents are considered to feel better-off when learning positively about themselves. We account for these motivational biases by assuming that workers are likely to process bad signals about their abilities as if they were good signals. We assume that agents update the actual processed information using Bayesian inference. Our updating process belongs to a wider class of learning processes defined by Rabin (2002) as quasi-Bayesian.[1]

In our framework, positive self-image arises because of individuals' mental processes that modify beliefs about abilities. Positive self-image can also be generated without assuming that individuals have a need to protect their ego by distorting their perceptions of successes and failures. For example, Van den Steen (2004) develop a model in which agents are assumed to have different priors about the probability of success on a

[1] "A person is modeled as having a specific form of misreading of the world meant to correspond to a heuristic error, but then is assumed to operate as a Bayesian given this misreading." Rabin (2002).

given task. In that context, individuals will tend to select actions for which they overestimate the likelihood of success. As a result, individuals will overvalue the probability with which their failures are due to bad luck rather than to insufficient talent. Another possibility is to generate positive self-image by considering that individuals have subjective perceptions of the levels of ability (Santos-Pinto and Sobel 2005). According to Santos-Pinto and Sobel, individuals develop positive self-image since they invest in improving the skills that are most relevant for their personal definition of abilities. In this paper we do not provide a theory of positive illusions, rather, we analyze how positive illusions affect cooperation in the workplace. In that respect, the specific process used to generate positive illusions is not of decisive relevance in our work.

Various researchers have studied the role of behavioral factors in the context of teams. They have focused on finding possible solutions to free riding arising in teams when efforts of its members are not observable.[2] Rotemberg (1994) demonstrates how altruism can improve workers' cooperation and welfare when complementarities exist among team members. Kandel and Lazear (1992) show how peer pressure can increase cooperation among workers by stressing how workers can reduce the negative effects of peer pressure by exerting higher levels of efforts. Gervais and Goldstein (2006) find that workers' biased self-perception facilitates cooperation among agents. The argument is that an overconfident agent overestimates his marginal product of effort leading himself and his coworker to exert more effort in the team. The authors show that both the self-confident and the rational workers can benefit from overconfidence.

Our framework differs from the ones previously mentioned since it eliminates free riding issues by assuming observability of coworkers' actions. We consider the most

[2] Free riding issues in teams have been studied in numerous papers such as Holmstrom (1982), Itoh (1991) or Che and Yoo (2001).

favorable case for workers' cooperation by focusing on teams with a sufficiently close level of collaboration such that agents are able to observe each others' performances and actions. In contrast to Gervais and Goldstein (2006), we find that workers' biased self-perception has a negative impact on cooperation in the workplace. Our approach is different from these authors since we analyze cooperation as the decision of team formation whereas they analyze the level of effort undertaken by agents assuming that a team has already been formed.

Other recent works examine team issues in the absence of moral hazard. Eaton and Hollis (2003), for example, consider teamwork when agents have private information about their own opportunities. They show that asymmetry of information among team members leads to insufficient teamwork and they justify the use of incentive schemes that over-reward joint work. Ishida (2005) focuses on information asymmetry between the principal and team members. In a repeated setting, the author justifies the use of low-powered incentives for individual work. Under such incentive schemes the principal creates a motive for workers to build a reputation for being cooperative and this facilitates peer monitoring. In our model, no asymmetry of information is present exante since performances are observable by both workers. Asymmetry of information arises as a consequence of self-serving biases. This is the case since learning biases imply that workers learn differently about their ability and the ability of their partner.

Our approach also differs from models involving incomplete contracts. For example, Hart and Moore (2006) distinguish between legally enforceable (consummate) performance and not legally enforceable (perfunctory) performance. They find in that context that rigid contracts may be preferred to flexible contracts as they tend to limit aggrievement and retaliation between buyers and sellers. This is the case because rigid contracts that fix the price of a good exante will work as a reference point for individuals' feelings of entitlements and this will reduce the possibility of disagreements among buyers and sellers. In our framework the only element that is not contractible is the unconscious mental process used by workers in order to build a positive self-image. In our model, if this mental process were contractible then workers could not develop a positive view of themselves. We do not need to assume that performances are not contractible in order to show that rigid contracts can be preferred to contracts that are contingent on workers' achievements.

To analyze team formation we consider a two-period model in which workers jointly decide whether to form a team or work alone. We assume that workers' abilities are unknown, and agents update their beliefs about abilities after receiving a signal at the end of the first period. We show that when workers suffer from self-serving attribution, cooperation among agents is undermined whatever the allocation rule that is considered for the group outcome. The negative impact of self-serving biases on team formation is referred to as the $teams\ inefficiency\ result$. We show that this result is robust to high degrees of workers' sophistication. Our model establishes a basic framework to analyze the necessary psychological conditions for individuals to form teams. We also analyze the design of teams contracts that foster team formation among self-serving workers. We show that fixed allocation rules can be preferred to contingent contracts based on coworkers' abilities. This finding is seen as the evidence of psychological limitations on contracting. As psychological biases are present, workers interpret information differently and this renders difficult the use of complex contracts that are contingent on workers' performances. We then apply our model to organizations with a particular focus on research institutions. We provide a psychological explanation for the fact that research institutions over-reward joint works.

The rest of this paper is organized as follows. We present and solve our model in the case of rational coworkers in Section 2 and analyze the model with self-serving biases in the third section. In Section 4, we study the robustness of the *teams inefficiency result* to the case of sophisticated workers willing to overcome their biases. Subsequently, we discuss applications of our model in Section 5. Section 6 concludes. All proofs are available in the appendix.

2 The benchmark model of team formation

In this section, we analyze the benchmark framework in which workers are assumed to behave as Bayesian inferers.

2.1 The team formation framework

We consider the case of two workers deciding whether to complete an individual or a team project. Examples of such decisions are found in the academia when researchers decide whether to write a single-authored or a coauthored paper. Workers may also be confronted with decisions to form teams in their organizations as in the case of the Koret Corporation described by Hamilton (2003). We propose to model team formation in a two-period game described as follows. At $t = 0$, the two coworkers decide simultaneously whether to undertake the individual or the group project. The team project is undertaken only if both workers agree to do so. At the end of the first period the outcome of the project chosen at $t = 0$ is observed by both workers. At $t = 1$, agents decide whether to continue with the project undertaken in the first period. The outcome associated to the project performed in the second period is observed at $t = 2$. Team members do not know neither their own ability to undertake the task nor the ability of their coworker. Workers update their beliefs about abilities at the end of the first period after observing the outcome of the project chosen in the first period. We assume agents are risk neutral so that they select their projects by maximizing expected payoffs. An agent $i \in \{1;2\}$ when working alone undertakes a project that is a success [failure] with probability q_i $[1-q_i]$ and delivers a payoff $X_{i,t} \equiv G$ $(B < G)$, where q_i is defined as Worker i's ability. The subscript t corresponds to time where $t \in \{0;1;2\}$. We drop the time subscript when not necessary. We assume a Beta prior distribution for individual abilities: $q_i \sim Beta(\alpha, \beta)$ and we denote $q^* = \frac{\alpha}{\alpha+\beta}$ the mean of this distribution.[3-4] The outcomes of the two individual projects are assumed to be independent. If workers choose to form a team, they are involved in a project that delivers the following payoff $\gamma\left(X_{1,t} + X_{2,t}\right)$, $\forall t \in \{1;2\}$. The total outcome of the group project is shared according to an allocation rule $\eta \in [0,1]$ so that Workers 1 and 2 get respectively payoffs $\eta\gamma\left(X_{1,t} + X_{2,t}\right)$ and $(1-\eta)\gamma\left(X_{1,t} + X_{2,t}\right)$. The parameter γ represents synergies obtained for working in a team. We assume γ is known by workers at $t = 0$. The absence of synergies corresponds to $\gamma = 1$. In that case the total outcome of the team project is the sum of the individual projects outcomes. In addition, we assume the existence of a learning by doing effect such that if workers repeat a project (a team or an individual project) the expected payoffs associated to that project are multiplied by $\phi \geq 1$. We consider no discount factors; the effect of discounting would be to reduce the role of learning about workers' abilities at $t = 1$.[5] The sequence of decisions as well as the payoffs of the individual and team projects are represented in Figure 1, where q^*

[3] The beta prior assumption is convenient since the beta distribution is a conjugate prior for the binomial problem considered here (Box and Tiao 1973). In addition, beta distributions can approximate any reasonably smooth unimodal distribution on $[0,1]$ (Lee 1997).

[4] Similar results are obtained if we consider workers with different prior abilities.

[5] A low discount factor would not be consistent with our aim since we want to consider projects for which learning and then self-attribution biases matter.

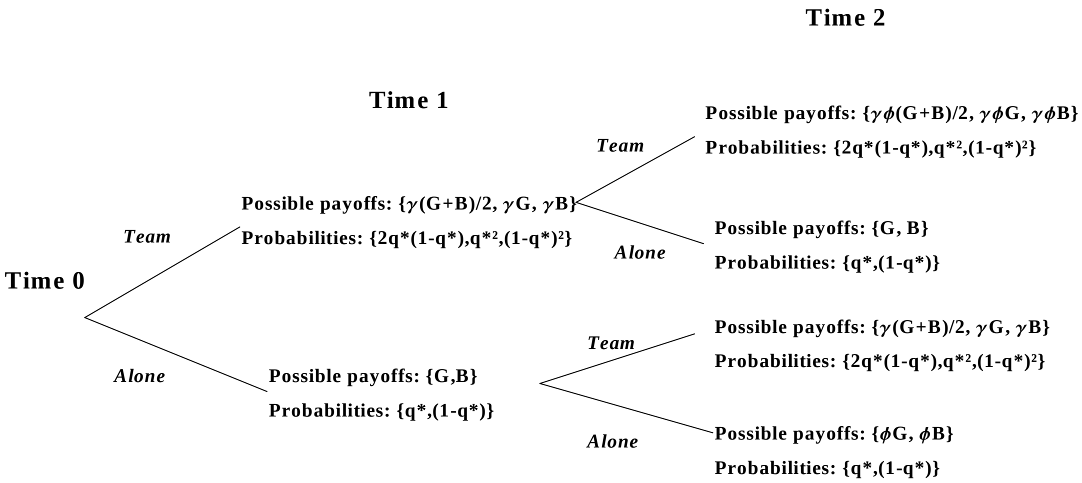

Figure 1: Timeline and payoffs for $\eta = \frac{1}{2}$

stands for the prior expected ability of workers.

2.2 Comments on the assumptions

Instead of assuming perfect observability of coworkers' performances, it may appear more natural to consider that workers learn more about their partner when they work as a team. We can study the case in which workers are able to observe others' performances only when they form a team. This leads to a framework in which workers may decide to hide bad news about their abilities in order to signal themselves as being high-ability coworkers. In this setting Propositions 1 and 2 are not modified for any of the equilibria of the game. This is the case since the conditions for team formation at $t = 0$ crucially depend on the conditions for team formation at $t = 1$ when a team has been formed at $t = 0$. Since these conditions do not change with respect to the benchmark model, the conditions for team formation at $t = 0$ are not modified.[6]

[6] The analysis of this game is available in an extended version of this paper that is available upon request.

Concerning the risk neutrality assumption, we have to mention that taking into account risk aversion is likely to strengthen our results. The idea is that, as self-serving biases increase, the uncertainty about team continuation at $t = 1$ rises. As a result, the negative impact of self-serving attribution on workers' cooperation is likely to be higher for risk averse agents.

Instead of assuming the presence of a learning by doing effect (ϕ), we can consider a fixed cost $C > 0$ incurred for shifting from the individual [team] project to the team [individual] project at $t = 1$.[7] We show that the main results of our paper are not modified.[8]

We consider in our model a situation in which workers have the possibility to leave the team at $t = 1$. However, there exist cases in which agents may attempt to commit at $t = 0$ to continue with the project started in the first period. We have to stress that commitment at $t = 0$ may be broken at $t = 1$ by one of the two workers. In our framework commitment is not credible as it happens in many real life situations in which an exante agreement can be broken without further costs. We consider such examples in the case of the academic profession in Section 5.[9]

2.3 Analysis of team formation

We consider an allocation rule under which the share of the group outcome obtained by an individual is equal to his relative ability. The relative ability of Worker i is defined as $\frac{\hat{q}_{i,t}}{\hat{q}_{i,t}+\hat{q}_{j,t}}$, $i \neq j, \forall (i,j) \in \{1;2\}, \forall t \in \{0;1\}$. We denote $\hat{q}_{i,t}$ the level of ability of Worker i as updated by a Bayesian inferer given information up to time t. Under this allocation

[7] You may assume that there is a fixed cost to pay for undertaking a project. You do not have to pay such a cost if you repeat the same type of project in the second period.

[8] We analyze this alternative specification of the model in an extended version of this paper.

[9] Two coauthors may not be able to credibly commit to continue working together since they know that one of the researchers can break the agreement at a low cost.

rule, Worker i's expected payoffs for a team project undertaken for the first time is $\gamma \hat{q}_{i,t}$. The next proposition shows that, in this case, workers form teams at $t = 0$ whenever $\gamma \geq 1$.[10] This result still holds if coworkers' prior abilities are different as long as both workers agree on the priors.

Proposition 1 *Under the relative ability allocation rule and in the absence of self-serving biases, teams are formed at* $t = 0$ *whenever* $\gamma \geq 1$.

Our proposition shows that by selecting a splitting rule that depends on updated workers' ability, the maximum level of workers' cooperation is attained.[11] As a result, we show that the efficient teams outcome (ETO) is attainable in the absence of self-serving biases, where the ETO corresponds to the payoffs obtained by team members when teams are formed at $t = 0$ and continued at $t = 1$ whenever $\gamma \geq 1$. We call efficient teams equilibrium (ETE) an equilibrium that implements the ETO.

3 The model with self-serving biases: first evidence of the teams inefficiency result

3.1 Assumptions on self-attribution biases

In this section we consider that workers suffer from biases in their learning process. Self-serving attribution as it is mentioned in the introduction can be seen as Bayesian learning with imperfect processing of negative signals. Researchers have found that

[10] In addition to the subgame-perfect Nash equilibrium considered in Proposition 1, the other subgame-perfect equilibria are as follows. By backward induction we obtain the following equilibria. 1) No workers form teams at $t = 1$ and teams are formed for $\gamma \geq \phi$ at $t = 0$. 2) No workers form teams at $t = 0$ and at $t = 1$. 3) Teams are formed at $t = 1$ and no teams are formed at $t = 0$. These equilibria involve weakly dominated strategies. In addition they involve strategies that prevent any cooperation in at least one of the two periods.

[11] However, one may think of real life examples in which workers are not paid with respect to their relative abilities. In the case of Economics research, credits are shared equally among team members.

positive personality information is efficiently processed whereas negative personality information is poorly processed (Kuiper and Derry 1982, Kuiper and McDonald 1982, Kuiper et al. 1985). We introduce inference biases by assuming that, with probability p, workers process bad signals about their ability as they were good signals. Our assumption implies a different treatment of bad and good signals. This asymmetry in the learning process is what we call biased self-attribution or self-serving learning. Workers are tempted to distort bad signals about their abilities in order to build a positive self-image. Through time, above average effects arise leading workers to see themselves as more talented than their coworkers. The latter effects generate a dispersion in coworkers' beliefs about their own ability and the ability of their coworker. Differences in perceptions about abilities will lead agents to break teams. The learning process that is considered in this section is described in Assumption 1. Workers are assumed to suffer from self-serving biases by mistakenly interpreting bad signals about their abilities.[12] According to Assumption 1, workers are unaware of their incentives to exhibit self-serving biases. Indeed, workers do not discount good news about their ability using the fact that such news can be bad news that have been distorted as a result of their self-serving tendencies. We assume that self-serving learning occur whether agents work alone or in a team.

Assumption 1 (Self-serving Learning)

We denote σ_{ij} Worker $i^{'}$s perception of Worker $j^{'}$s performance at $t = 1, \forall\, (i, j) \in \{1; 2\}^2$. We assume that, with probability p, a Worker i perceives his bad performance at $t = 1$ $(X_{i,1} = B)$ as if it was a good performance $(\sigma_{ii} = G)\,, \forall i \in$

[12]Learning biases can be modeled as a result of errors in information processing or as memory imperfections. However, this distinction between the different origins of learning biases is not central to our results and to their implications.

$\{1;2\}$.[13] The updating rule at $t=1$ is described as follows, $\forall i \in \{1;2\}$.

$E_{i,S}[q_i \mid \sigma_{ii}=x] \equiv E[q_i \mid X_{i,1}=x]\,,\ \forall x \in \{B;G\}$

A worker updates his coworker's ability using Bayesian inference and correct information processing, that is $\sigma_{ij}=X_{j,1}$, $\forall i \neq j$ and $(i,j) \in \{1;2\}^2$.

We denote $E_{i,S}$ the expectation of workers suffering from self-serving biases p.[14] We introduce a subscript i for the expectation of Worker i since when learning biases are present coworkers' expectations may not coincide. We assume that the two coworkers suffer from learning biases. According to our learning process, workers but they are assumed to update the beliefs about their own ability differently from the beliefs about others' abilities. Agents are Bayesian inferers when updating others' abilities but assumed to suffer from self-serving biases when updating their own ability. There is evidence in the literature in Psychology that individuals see themselves more positively than others see them. For example, Lewinsohn et al. (1980) compared the ratings made by observers and by college students themselves about personality characteristics like friendliness, warmth and assertiveness of students involved in a group interaction task. They found that self-ratings were significantly more positive than observers' ratings. We consider the case in which team workers do not suffer from learning biases in assessing their coworker's ability. However, it is possible to argue that workers forming a team together assess positively the ability of their coworker. The results of this section still hold as long as workers' self-serving biases in assessing their ability are more pronounced than in estimating the ability of their coworker.

[13] Workers' biases are assumed to be independent.

[14] Alternatively, we can consider the case of two agents with different degrees of self-serving attribution: $p_1 \neq p_2$. The results derived below continue to hold taking $p \equiv Max\{p_1;p_2\}$.

Another important element of our quasi-Bayesian learning process model is the degree with which workers are aware of their biases. Assumption 1 is consistent with the fact that workers are not fully aware of their biases (Assumptions $2a$ and $2b$). A span of assumptions on workers' awareness of self-serving biases is considered below. In the rest of the paper, we assume that the two coworkers have the same degree of awareness of biases.

Assumption 2 (Awareness of Biases)

Assumption 2a: (the naive case). Workers are unaware of their learning biases as well as of their coworker's biases.

Assumption 2b: (asymmetric awareness of biases). Workers are unaware of their learning biases but are aware of their coworker's biases.

Assumption 2c: (full awareness of biases). Workers are aware of their coworker's biases as well as their own biases.

A first possibility is to consider that workers are naive inferers that are not aware of their own biases (Assumption $2a$). In this case the effect of learning biases will not translate to the formation of teams at $t = 0$. As many teams will be formed at $t = 0$ in the self-serving learning and in the Bayesian learning cases but more teams will be split at $t = 1$ when self-serving biases are present. An alternative assumption is to consider asymmetry in the level of awareness of biases (Assumption $2b$). Workers may recognize that others learn positively about themselves but may not be able to identify their own biases (Pronin, Lin and Ross 2002). Under Assumption $2c$ workers are aware of their own biases; this assumption is used in Bénabou and Tirole (2002).[15] We consider

[15] Bénabou and Tirole refer to this assumption as metacognition.

this assumption in Section 4. As we will establish in this paper, our result on the negative effects of self-serving biases on team formation (*teams inefficiency result*) is robust to different degrees of sophistication of workers. Psychologists have stressed the limited awareness of agents about their mental processes (Epstein 1983, Gilbert et al. 1998). This view is consistent with Assumptions $2a$ and $2b$. According to Assumptions $2a$ and $2b$, the conditions for team formation at $t = 1$ are modified compared to the benchmark model if at least one coworker suffers from self-serving biases, this occurs with probability $\Pi = 2p - p^2$.[16]

3.2 Analysis of team formation in the presence of self-serving biases

We present the intuition of our *teams inefficiency result* by showing that it may be impossible to design an allocation rule such that workers may perceive themselves better-off working in a team whenever $\gamma \geq 1$. As a result, we show that the efficient teams outcome (ETO) is not attainable in the presence of self-serving biases.

Consider for a moment that both workers performed badly in the first period, so that $(X_{1,1}, X_{2,1}) = (B, B)$. We also assume that both workers suffer from self-serving biases at $t = 1$. This occurs a priori with probability $\frac{\beta^2}{(\alpha+\beta)^2}p^2$. Denoting $\hat{q}_{i,t}^j$ the ability of agent i estimated by agent j at time t, we obtain by assumptions that $\hat{q}_{1,1}^1 = \hat{q}_{2,1}^2 = \frac{\alpha+1}{\alpha+\beta+1}$ and $\hat{q}_{2,1}^1 = \hat{q}_{1,1}^2 = \frac{\alpha}{\alpha+\beta+1}$.[17] Considering that a team has been formed at $t = 0$ and denoting η_i the share of the group outcome obtained by Worker $i \in \{1; 2\}$ ($\eta_1 + \eta_2 = 1$), a team will be formed at $t = 1$ if both workers are better-off working as a team. This occurs if the following conditions are satisfied:

[16] Workers' self-serving biases are assumed to be independent.

[17] We take $G = 1$ and $B = 0$ without loss of generality.

$$\begin{cases} \gamma\phi\eta_1\left(\hat{q}^1_{1,1}+\hat{q}^1_{2,1}\right)\geq\hat{q}^1_{1,1} \\ \gamma\phi\left(1-\eta_1\right)\left(\hat{q}^2_{1,1}+\hat{q}^2_{2,1}\right)\geq\hat{q}^2_{2,1} \end{cases}$$

$$\Leftrightarrow \gamma\phi \geq \frac{\frac{\alpha+1}{2\alpha+1}}{Min\{\eta_1;1-\eta_1\}} > 1 \text{ since } \underset{\eta_1}{Min}\frac{\frac{\alpha+1}{2\alpha+1}}{Min\{\eta_1;1-\eta_1\}} = \frac{2\alpha+2}{2\alpha+1} > 1.$$

As a result, whatever the sharing rule η_1, the condition for team formation is more demanding in terms of synergies than in an ETE since $\frac{\frac{\alpha+\beta+1}{2(\alpha+\beta)+1}}{Min\{\eta_1;1-\eta_1\}} > 1, \forall\ 0 \leq \eta_1 \leq 1$. Recall that we refer to the ETO as the situation in which workers are willing to form teams at $t = 0$ whenever $\gamma \geq 1$. We establish a more general result in Proposition 2.

Proposition 2 *Whatever the allocation rule considered, as long as at least one coworker suffers from self-serving biases, the* ETO *cannot be attained.*

Self-serving biases create a divergence in beliefs among coworkers. A direct consequence is that even when allocation rules are flexible and $\gamma \geq 1$, an allocation rule permitting team formation may not exist. In the appendix, we show that the *teams inefficiency result* captured in Proposition 2 is robust to general synergy functions.

Our result about the impossibility to find an allocation rule that ensures a sufficiently high level of cooperation is in line with the experimental results of Babcock et al. (1995) and Babcock and Loewenstein (1997) stating that self-serving biases tend to prevent defendants and plaintiffs from reaching an agreement about a settlement. The next corollary states that self-serving biases have a negative effect on aggregate expected welfare.[18]

Corollary 1 *Whatever the allocation rule* (η_1)*, the aggregate expected welfare in the case of self-serving learning is at most as high as in the case of Bayesian workers using a relative ability allocation rule.*

[18] The aggregate welfare is the sum of the outcomes obtained by the two workers.

We have shown that when self-serving biases are present, it is impossible to obtain the maximum level of workers' cooperation that consists in forming teams whenever $\gamma \geq 1$. This is so because biased self-attribution implies a divergence in beliefs among agents. Conflicting beliefs prevent workers from agreeing on the relative ability sharing rule that would make both agents better-off by attaining a higher level of cooperation. The intuition is that cooperation is undermined either by a rigidity in allocation rules as for example the equal splitting rule (Farrell and Scotchmer 1988) or by a "rigidity in beliefs". Rigidity in beliefs arises when self-serving biases are present since then workers stick to excessively positive beliefs about themselves being convinced that they hold the truth.

As a corollary of Proposition 2, we show that there exist no long term commitment contracts implementing the ETO. To do so we define a contract as the share of the group outcome η_i distributed to Worker i at $t = 1, \forall i \in \{1; 2; 3\}$. We consider budget balanced contracts ($\eta_1 + \eta_2 = 1$) as well as contracts involving a third party ($\eta_1 + \eta_2 + \eta_3 = 1$).[19] We consider the case of a third party that can punish team breaks by imposing a penalty on the agent(s) splitting the team at $t = 1$. We assume that the third party is not involved in team production. However, he is assumed to be able to observe without biases the performances of the two coworkers. We use the following definitions.

Definition 1 *A* long term commitment contract *is such that it can be renegotiated at $t = 1$ if all parties agree to do so.*[20]

Definition 2 *A contract is* renegotiation-proof *if it is impossible to design at $t = 1$ a new contract that increases the utility of one agent without reducing the utility of the*

[19] The group outcome is distributed in its totality to workers. This definition is similar to the one used in Bartling and von Siemens (2004).

[20] *This definition is taken from Salanié (1997).*

other agent.

By definition, a long term commitment contract is renegotiation-proof. Corollary 2 can then be derived from Proposition 2.

Corollary 2 *In the model with two agents and in the case of a third party, there exist no long term commitment contracts that can implement the ETO.*

In Corollary 2, we extend the result of the initial framework to the presence of a third agent. A third party may not be able to implement teamwork by imposing penalties on agents breaking the teams. Indeed, the third agent is exposed at $t = 1$ to contracts offers from his two coworkers that give him higher expected payoffs when teams are not formed. As a result, the implementation of the ETO is impossible when renegotiation is possible.

4 The teams inefficiency result with sophisticated agents

4.1 Assumptions: the revelation game

In this section agents are assumed to be aware of their incentives to process information with biases (Assumption 2c). As it is considered in the previous sections, workers update others' abilities using Bayesian inference. We take p_G $[p_B]$ to be the expected probability given information at $t = 0$ that $X_{i,1} = G$ $[X_{i,1} = B]$, $\forall i \in \{1; 2\}$.

Assumption 3 (Self-serving Learning for sophisticated workers)

We denote σ_{ij} Worker i's perception of Worker j's performance at $t = 1, \forall (i, j) \in \{1; 2\}^2$. We assume that, with probability p, a Worker i perceives his bad performance at $t = 1$ $(X_{i,1} = B)$ as if it was a good performance $(\sigma_{ii} = G)$, $\forall i \in$

$\{1;2\}$.[21] The updating rule at $t=1$ is described as follows, $\forall i \in \{1;2\}$.

$$\left[\begin{array}{l} E_{i,S}\left[q_i \mid \sigma_{ii}=B\right] \equiv q_B \equiv E\left[q_i \mid X_{i,1}=B\right] \\ E_{i,S}\left[q_i \mid \sigma_{ii}=G\right] \equiv \hat{q}_G \equiv \frac{pp_B}{pp_B+p_G} E\left[q_i \mid X_{i,1}=B\right] + \frac{p_G}{pp_B+p_G} E\left[q_i \mid X_{i,1}=G\right] \end{array}\right.$$

A worker updates his coworker's ability using Bayesian inference and correct information processing, that is $\sigma_{ij} = X_{j,1}$, $\forall i \neq j$ and $(i,j) \in \{1,2\}^2$.

The self-serving learning process is assumed to be common knowledge. We consider that workers are aware of their incentives to be biased. Workers will try to overcome their biases by recovering the correct signals about their abilities. The inefficiency result captured in Proposition 2 is based on the assumption that workers are unable to recover information about their own ability. This behavior is consistent with Assumptions $2a$ and $2b$. In this section, we assess the robustness of the $teams\ inefficiency\ result$ by considering sophisticated agents of the type described in Bénabou and Tirole (2002).

We define a contract as the share of the group outcome η_i distributed to Worker i at $t=1, \forall i \in \{1;2\}$.[22] The set of contracts analyzed are budget balanced, that is the group outcome is distributed in its totality to workers $(\eta_1 + \eta_2 = 1)$. We consider contracts that can be contingent on coworkers' performances received at $t=1$. The difficulty is that workers' suffering from self-serving biases may disagree about the signals received at $t=1$. To tackle this issue we consider that contracts are contingent on the signals revealed by the agents rather than on the signals effectively observed. We modify the initial framework by introducing a revelation game at $t=1$ after workers have observed their performances on the first period project (Figure 2).[23] Workers are interested

[21] Workers' biases are assumed to be independent.

[22] The share of the group outcome given to Worker 1 in the first period is not considered further since $\eta_1 = \frac{1}{2}$ ensures team formation at $t=0$ if team formation is obtained at $t=1$.

[23] In the context of sophisticated workers, we assume that performances are not verifiable by the court. If performances were verifiable by the court, workers could reach the ETO by asking the court to reveal workers' performances. Evidently, such a process can be costly to workers.

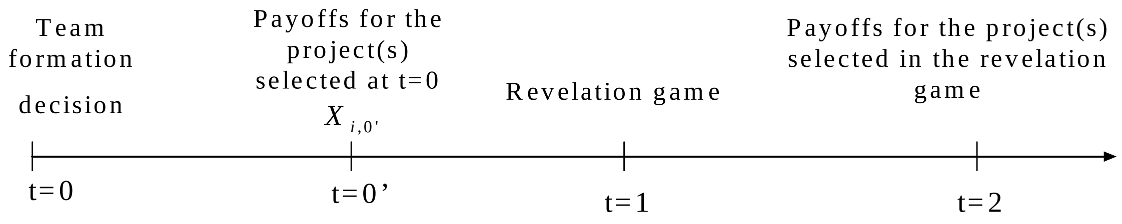

Figure 2: Timeline for the model with sophisticated agents and a revelation game.

in communicating about their perceived abilities since they know that their coworker is an objective observer of their performances. On aggregate workers have complete information about abilities since Worker 1 [2] knows Worker 2 [1] ability level at $t = 1$.

The structure of the revelation game played at $t = 1$ is as follows.

At $t = 1$ each coworker chooses an action $a_i \equiv (a_{i1}, a_{i2})$ $\forall i \in \{1, 2\}$, where a_i is a vector of messages that belongs to the set S of possible signals observed at $t = 0^{'}$. The set S is actually the set of possible types of coworkers. This is the case since the perception of performances by the agents constitutes their private information.[24]

At $t = 1'$ where $1' \in]1, 2[$, workers decide either to continue with the project selected at $t = 0$ or to undertake the other project. We denote $b_i \in B \equiv \{T; NT\}$, Worker i's action at $t = 1'$, $\forall i \in \{1; 2\}$, where T $[NT]$ stands for forming a team [working alone].

The actions of the two agents will determine the share of the group outcome given to the first coworker (η) as a function of the revealed signals, that is $\eta \equiv \eta_1 (a_{11}, a_{12}, a_{21}, a_{22})$.[25] We denote $V_i (a_i, a_j, b_i, b_j)$ the expected payoffs obtained by Worker i when undertaking the second period project, $\forall i \neq j$ and $(i, j) \in \{1; 2\}^2$.

[24] The set of possible messages being the set of types, we can use the Revelation Principle and conclude that our results continue to hold for any message space. The Revelation Principle can be applied to our model since it can be represented as a normal form game of a static Bayesian game.

[25] We denote $\eta_i (a_{11}, a_{12}, a_{21}, a_{22})$ the share of the group outcome obtained by Worker i.

Given that workers assess each others' abilities as Bayesian inferers, we may wonder if allowing workers to communicate will lead agents to eliminate their learning biases and cooperate efficiently.[26] The result captured in Proposition 3 shows that such conjecture is not verified, an ETE being impossible to achieve.

Definition 3 *Under Assumption* 3*, a* PBE *of the revelation game is* $A^* \equiv (a_1^*, a_2^*, b_1^*, b_2^*)$ *that solves* (2) *and* (3)*:*

(2) $\max\limits_{a_i \in S} V_i \left(a_i, a_j^*, b_1^*, b_2^*\right), \forall i \neq j,\ (i,j) \in \{1;2\}^2.$

(3) $\max\limits_{b_i \in B} V_i \left(a_1^*, a_2^*, b_i, b_j^*\right), \forall i \neq j,\ (i,j) \in \{1;2\}^2.$

Where $V_i \equiv 1_{NT} E_{i,S} \left[q_i | \ \sigma_{ii}, a_j\right] + 1_T E_{i,S} \left[\eta_i \left(a_i, a_j\right) \left(q_i + q_j\right) \mid X_{j,1,} \sigma_{ii}, a_j\right]$

We denote 1_T $[1_{NT}]$ *the indicator function that takes value one for* $b_1 = b_2 = T$ $[(b_1, b_2) \neq (T, T)]$.

A PBE is defined for a given contract function $\eta : x \mapsto \eta(x)$, where $x \in \{B; G\}^4$ and $\eta(x) \in [0, 1]$.

4.2 The teams inefficiency result

Proposition 3 is the counterpart of Proposition 2 when agents are learning about their biases and have the possibility to communicate about their perceived performances through a revelation game.

Proposition 3 *There exist no Perfect Bayesian Equilibria* (PBE) *that implement the* ETO*.*

Proposition 3 shows that the *teams inefficiency result* first stated in Proposition 2 is robust to the case of sophisticated workers that attempt to overcome their

[26] This would be the case when teams are formed whenever $\gamma \geq 1$.

biases. Workers are unable to reach the ETO because they have an incentive to reveal themselves as being high-ability workers in order to obtain a higher share of the group outcome. These incentives to lie implies that truthful telling is costly to achieve. Indeed, workers tell the truth in equilibrium only if the allocation rule of the group outcome is a fixed rule that is not contingent on (a_1, a_2), i.e. $\eta(a_1, a_2) = \bar{\eta}$. However, fixed allocation rules do not provide the adequate incentives for workers to form teams since then high-performance workers will perceive their team rewards as being insufficient. In the case of fixed allocation rules the ETO is not attainable even in the presence of complete information.[27]

4.3 Team contracts and workers' cooperation

In the next proposition we establish, using the revelation game presented in the last subsection, the conditions under which uninformative and truthful telling PBE lead to team formation. Uninformative PBE are pooling equilibrium in which every type of worker plays the same strategy. Under the pooling equilibrium, no information about workers' biases is revealed implying that agents are unable to reduce their learning errors. Under the uninformative PBE the revelation game is of no use to improve cooperation among workers. Truthful telling equilibria (TTE) are such that all the information is revealed in equilibrium. This occurs if workers truthfully reveal their perceived performances $(a_i = \sigma_i \equiv (\sigma_{ii}, \sigma_{ij}), i \neq j)$. In that case, workers are able to recover the true information about their performances and this implies that learning biases are fully recognized.

Proposition 4 *i) In the case of non-contingent allocation rules ($\eta(a_1, a_2) = \eta$), there*

[27] The proof of this result is trivial and is available upon request.

exists an uninformative PBE that always leads to team formation at $t = 1$ *for* $\gamma\phi \geq M$ $\left[\frac{\gamma}{\phi} \geq M\right]$ *when a team has [not] been formed at* $t = 0$*, where* $M \equiv Max\left\{\frac{\hat{q}_G}{\eta(q_B+\hat{q}_G)}; \frac{\hat{q}_G}{(1-\eta)(q_B+\hat{q}_G)}\right\}$ *and* $\hat{q}_G = wq_G + (1-w)\,q_B$ *with* $w = \frac{p_G}{p_G+pp_B}$.

ii) In the case of non-contingent allocation rules, there exists a truthful telling PBE that always leads to team formation at $t = 1$ *for* $\gamma\phi \geq L > M$ $\left[\frac{\gamma}{\phi} \geq L\right]$*, where* $L \equiv Max\left\{\frac{q_G}{\eta(q_B+q_G)}; \frac{q_G}{(1-\eta)(q_B+q_G)}\right\}$.

A TTE implementing the ETO is attainable only if the allocation rules considered are independent of workers' types. In a TTE, team formation is obtained at best for $\gamma\phi \geq \frac{2q_G}{q_B+q_G} > \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2q_G}{q_B+q_G}\right]$ whereas team formation is achieved for $\gamma\phi \geq \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2\hat{q}_G}{q_B+\hat{q}_G}\right]$ in the uninformative equilibrium.[28-29] Assuming $\eta = \frac{1}{2}$, for $\gamma\phi \in \left[\frac{2\hat{q}_G}{q_B+\hat{q}_G}, \frac{2q_G}{q_B+q_G}\right[\equiv H$ $\left[\frac{\gamma}{\phi} \in H\right]$, truthful revelation leads to an equilibrium in which teams are formed only when the signals received at $t = 0'$ are symmetric whereas the uninformative PBE leads to team formation whatever the signals received. As a result, Proposition 4 implies that full revelation of information can lead to less team formation than in the case in which no information is revealed. This is the case since under the uninformative equilibrium, workers who perceive themselves as being good are aware of the possibility that they may have performed badly. As a result, if no information is revealed in equilibrium, workers who performed well in the first period think that their level of ability is $\hat{q}_G < q_G$. In that case, high-performance workers appear to be self-critical and then more inclined to form teams in the second period than in the case of fully informed workers. In particular for $\gamma\phi \in]L, M[$ uninformed high-performance workers will always form teams at $t = 1'$ even if they only receive half

[28]Notation: in the rest of the paper, the condition into brackets will correspond to the case in which a team has not been formed at $t = 0$.

[29]We consider the most favorable case for team formation by taking $\eta = \frac{1}{2}$.

of the group outcome whereas informed high-performance workers would refuse to work with low-performance workers in that case. Given that the uninformative equilibrium is attainable in the absence of a revelation game by fixing the sharing rule of the group outcome to $\eta = \frac{1}{2}$, workers may decide not to play the revelation game at $t = 1$. This result is established in the next corollary.

Corollary 3 *The revelation mechanism is accepted by coworkers as long as* $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} < \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$.

Proposition 4 and Corollary 3 show how the process of acquiring information about one's own biases can be costly in terms of cooperation. This stresses that ignoring the possibility that we are biased can be an optimal strategy.[30] This strategy would be justified in our case for sufficiently high levels of synergies $\gamma\phi \in H$ $\left[\frac{\gamma}{\phi} \in H\right]$ since then sophisticated workers that are willing to overcome their own biases by playing the revelation game will be worse-off than agents that decide to adhere to the equal sharing rule $\eta = \frac{1}{2}$ without playing the revelation game. However, for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} < \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$ and $\eta = \frac{1}{2}$, teams are formed when signals are symmetric (with probability $p_{BB} + p_{GG}$) in the TTE whereas teams are formed only when signals are symmetric and no biases have occurred (with probability $(1-p)^2 p_{BB} + p_{GG}$) in the absence of a revelation game.[31] Not being aware of one's own biases is logically more detrimental as p increases for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} < \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$. This is the case since as p rises the probability of team formation under the uninformative equilibrium ($(1-p)^2 p_{BB} + p_{GG}$) decreases. It appears that promoting communication among workers is beneficial in terms of cooperation for

[30] A behaviour that has been observed by psychologists, see Epstein (1983) and Gilbert et al. (1998).

[31] We use the following notations: $p_{kl} = E\left(P\left[X_{1,1} = k; X_{2,1} = l \mid I_0\right]\right)$ where $(k,l) \in S \equiv \{(B,B);(G,B);(B,G);(G,G)\}$ and I_0 is the information set at $t = 0$, that is the prior information on workers' abilities.

sufficiently small levels of synergies. For high levels of synergies contracts based on rigid allocation rules and no communication mechanisms may be preferred. Corollary 3 stresses how simple contracts can be preferred to more complex contracts. This is shown without assuming differences in costs of writing and implementing the contracts. Our interpretation of Corollary 3 is that there exist psychological limitations on contracting activities.[32]

In the next proposition we derive, assuming a team has been formed at $t = 0$, the contracts that are most likely to lead to team formation when $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$. These contracts are defined below and compared in Proposition 5. Similar contracts can be defined if a team has not been formed at $t = 0$ by substituting ϕ by $\frac{1}{\phi}$ in the definitions of contracts (C^2_{TTE}) and (C^3_{TTE}). For $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ we know that contracts based on fixed allocation rules cannot ensure team formation whenever $\gamma \geq 1$. We then consider contingent contracts and analyze the truthful telling PBE associated to these contracts. We use the set S^{θ} defined as follows:

$$S^{\theta} \equiv \left\{ \begin{array}{l} (q, r, q, r) \, \forall \, (q, r) \in S, \;\; (G, B, k, l) \, \forall \, (k, l) \in S \, \hat{A} \, (G, B) \\ (m, m, B, G) \, \forall m \in \{B, G\} \end{array} \right\}$$

Definition 4 *Contract $(C^1_{TTE,\bar{\eta}})$ is defined by the following system of equations:*

$$(C^1_{TTE,\bar{\eta}}) \Leftrightarrow \left\{ \eta\,(i, j, k, l) = \bar{\eta}, \forall\,(i, j, k, l) \in S^2 \right.$$

$C^1_{TTE,\bar{\eta}}$ is the contract associated to the TTE derived in Proposition 4. It is such that allocation rules are independent of the signals revealed by workers at $t = 1$. This contract leads to team formation for $\gamma\phi \geq \frac{2q_G}{q_B + q_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2q_G}{q_B + q_G}\right]$ for $\bar{\eta} = \frac{1}{2}$.

Definition 5 *Contract (C^2_{TTE}) is defined by the following conditions:*

[32] This cost arises endogenously. In the literature on contract theory and inequity aversion, a psychological cost is directly introduced in the utility function (Englmaier and Wambach 2006, Bartling and von Siemens 2004, Rey Biel 2005).

$$(C^2_{TTE}) \Leftrightarrow \begin{cases} \eta_{GBGB} \in \left[\frac{q_G}{\gamma\phi(q_B+q_G)}, 1-\frac{q_B}{\gamma\phi(q_B+q_G)}\right], \eta_{GBGG}=\eta_{GBGB} \\ (\eta_{GBBB}, \eta_{GBBG}, \eta_{BBBG}) \in A^3, \textit{ where } A \equiv \left[0, \frac{1}{2\gamma\phi}\right[\\ \eta_{BGBG} \in \left[\frac{q_B}{(q_B+q_G)\gamma\phi}, 1-\frac{q_G}{\gamma\phi(q_B+q_G)}\right], \eta_{GGBG}=\eta_{BGBG} \\ (\eta_{GGGG}, \eta_{BBBB}) \in B^2, \textit{ where } B \equiv \left[\frac{1}{2\gamma\phi}, 1-\frac{1}{2\gamma\phi}\right] \\ \forall (i,j,k,l) \notin S^{\theta}, \eta_{ijkl}=0 \end{cases}$$

Contract (C^2_{TTE}) is such that allocation rules depend on the signals revealed by coworkers at $t=1$. In particular, considering $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$, the share of the group outcome given to the first worker is higher [lower] than equal splitting since $\frac{q_G}{\gamma\phi(q_B+q_G)} > \frac{1}{2}$ $\left[1-\frac{q_G}{\gamma\phi(q_B+q_G)} < \frac{1}{2}\right]$ for $(X_{1,1}, X_{2,1}) = (G,B)$ $[(B,G)]$. This contingent contract associated with full revelation of information in equilibrium allows workers to be rewarded based on their true relative ability. However, this contract does not permit teams to be formed when both workers receive a bad signal and at least one of them suffers from self-serving biases. This is the case since truthful revelation is not a possible equilibrium when $\gamma\phi < \frac{q_B+3q_G}{2(q_B+q_G)}$ $\left[\frac{\gamma}{\phi} < \frac{q_B+3q_G}{2(q_B+q_G)}\right]$ if teams are formed for both $\sigma \equiv (\sigma_1, \sigma_2) \in \Sigma$ and $(X_{1,1}, X_{2,1}) \in V$. We denote $\Sigma \equiv \{(G,B,B,G); (G,B,B,B); (B,B,B,G)\}$ and $V \equiv \{(G,B); (B,G)\}$. In order to ensure team formation for $(X_{1,1}, X_{2,1}) \in V$, we have to prevent team formation for $\sigma \in \Sigma$ by taking η_{GBBG}, η_{GBBB} and η_{BBBG} sufficiently low, that is inferior to $\frac{1}{2\gamma\phi}$.

Definition 6 *Contract (C^3_{TTE}) is defined for $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$ as follows.*

$$(C^3_{TTE}) \Leftrightarrow \begin{cases} \eta_{GBGB} \in \left[\frac{q_G}{\gamma\phi(q_B+q_G)}, 1-\frac{1}{2\gamma\phi}\right] \\ \eta_{BGBG} \in \left[\frac{1}{2\gamma\phi}, 1-\frac{q_G}{\gamma\phi(q_B+q_G)}\right] \\ \eta_{GBGG}=\eta_{GBBB}=\eta_{GBGB}, \eta_{GGBG}=\eta_{BBBG}=\eta_{BGBG} \\ (\eta_{GGGG}, \eta_{BBBB}) \in B^2, \textit{ where } B \equiv \left[\frac{1}{2\gamma\phi}, 1-\frac{1}{2\gamma\phi}\right] \\ \eta_{GBBG} \in \left[0, \frac{1}{2\gamma\phi}\right[, \forall (i,j,k,l) \notin S^{\theta}, \eta_{ijkl}=0 \end{cases}$$

Contract (C^3_{TTE}) depends, similarly to contract (C^2_{TTE}), on the signals revealed by coworkers at $t = 1$. Contract (C^3_{TTE}) is defined for $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$ whereas contract (C^2_{TTE}) is implementable for any $\gamma \geq 1$. The reason is that for $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$, contract (C^2_{TTE}) can be improved by taking $\eta_{BBBG} = \eta_{BGBG}$ and $\eta_{GBBB} = \eta_{GBGB}$ since then teams can be formed for both $\sigma \in \{(G,B,B,B);(B,B,B,G)\}$ and $(X_{1,1}, X_{2,1}) \in \{(G,B);(B,G)\}$. However, contract (C^3_{TTE}) does not ensure team formation for any $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$ since teams are not formed when both workers receive a bad signal and both workers exhibit self-serving learning.

The three contracts previously defined do not strictly dominate each other, choosing the best contract depends on the level of synergies and on the level of learning biases. This result is stated in Proposition 5, where the Best contract is defined as the contract implementing the highest coworkers' expected aggregate welfare in equilibrium.

Proposition 5 *Contracts that lead to the highest coworkers' expected payoffs for* $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ *are as follows.*

i) For $\gamma\phi < \frac{q_B+3q_G}{2(q_B+q_G)}$ $\left[\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}\right]$ *and* $p_G < \frac{(2p-p^2)p_B}{2}$ $\left[p_G \geq \frac{p^2 p_B}{2}\right]$, $(C^1_{TTE,\frac{1}{2}})$ *is the Best contract.*

ii) For $\gamma\phi < \frac{q_B+3q_G}{2(q_B+q_G)}$ *and* $p_G \geq \frac{(2p-p^2)p_B}{2}$, (C^2_{TTE}) *is the Best contract.*

iii) For $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$ *and* $p_G \geq \frac{p^2 p_B}{2}$, (C^3_{TTE}) *is the Best contract.*

We know from Proposition 4 and the definitions above that contract $(C^1_{TTE,\frac{1}{2}})$ leads to individual work when the signals received are asymmetric whereas team formation is obtained in that case for the two other contracts. The three contracts are not equivalent since contracts (C^2_{TTE}) and (C^3_{TTE}) are preferred when self-serving biases are not too high, that is respectively when $p(2-p) \leq \frac{2p_G}{p_B}$ and $p \leq \sqrt{\frac{2p_G}{p_B}}$. An increase in coworkers' learning biases (p) does not affect the probability $(p_{GG} + p_{BB})$ with which a team is formed under contract $(C^1_{TTE,\frac{1}{2}})$ whereas it decreases the frequency with which teams are formed under contracts (C^2_{TTE}) and (C^3_{TTE}). In the next corollary, we derive from Propositions 4 and 5 the conditions under which contracts stating fixed allocation rules are dominated by contracts based on allocation rules that are contingent on coworkers' revealed signals.

Corollary 4 *i) For* $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ *and* $\gamma\phi < \frac{q_B+3q_G}{2(q_B+q_G)}$, *contingent allocation rules are strictly preferred to fixed allocation rules when* $p\,(2-p) \leq \frac{2p_G}{p_B}$.

ii) For $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ *and* $\gamma\phi \geq \frac{q_B+3q_G}{2(q_B+q_G)}$, *contingent allocation rules are strictly preferred to fixed allocation rules when* $p \leq \sqrt{\frac{2p_G}{p_B}}$.

iii) For $\gamma\phi \geq \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ *contingent allocation rules are strictly dominated by fixed allocation rules.*

Corollary 4 motivates the use of contingent allocation rules as long as synergies and biases are not too high. If synergies are high $\left(\gamma\phi \geq \frac{2\hat{q}_G}{q_B+\hat{q}_G}\right)$ then high-performance workers will be willing to form teams even if they are not rewarded according to their relative ability (*i.e.* $\eta = \frac{1}{2}$). In that case, fixed allocation rules implement the ETO. For $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ fixed allocation rules are preferred to contingent contracts when p is high (*i.e.* $p\,(2-p) > \frac{2p_G}{p_B}$ or $p > \sqrt{\frac{2p_G}{p_B}}$). This is the case since workers only have incentives to lie in the revelation game when allocation rules are contingent. And, the incentives to lie are stronger for high values of p since then workers' information is more asymmetric. Corollary 4 helps us understand why fixed allocation rules are commonly observed. We show that it may be due to the impossibility for agents to identify their learning errors from their coworker. Our corollary implies that partnerships based on fixed allocation rules can be justified when synergies are sufficiently high or when self-serving biases are prevalent.[33]

5 Applications

In this section, we first analyze cooperation in the academic profession. We then study the more general case of organizations with a special focus on the differences between Japanese and US firms.

5.1 Coauthorship

5.1.1 The issue of cooperation among researchers

There exists evidence that research institutions and academic departments over-reward joint works. McDowell and Smith (1992) reject the hypothesis according to which the weighting of coauthored articles in the promotion of academics is equal to the inverse of the number of coauthors. The authors use answers to a questionnaire given to 378 Economics researchers in 20 of the most important US institutions from 1968 to 1975. They find that each author of a paper with n coauthors is rewarded more than $\frac{1}{n}$ of the credits given to a single authored paper of the same quality. Schinski, Kugler and Wick (1998) using a survey of 140 Finance academics in 1995 confirm the results of McDowell and Smith (1992). Similar conclusions have been obtained by Liebowitz and Palmer (1983) and by Long and McGinnis (1982) in other fields than Economics. The rules used by research institutions can be interpreted as a mechanism to provide incentives for researchers to write joint papers. This may reveal a concern for insufficient coauthorship among researchers.

[33] Such conditions may be satisfied for consulting partnerships if we take into account that experts tend to be more self-confident than non-experts (Griffin and Tversky 1992).

5.1.2 An explanation based on self-serving biases

As a direct application of our model, we consider the decision of two researchers to write a joint paper. Researchers, similarly to other individuals, appear to suffer from self-serving tendencies. In particular, as it is found in Caruso, Epley and Bazerman (2005a, 2005b), researchers are inclined to overestimate their contribution to joint papers. When coauthors were asked to assess their relative amount of work in the team output (in percentage terms), the sum of the estimated contributions typically summed up to a number higher than one.[34] If self-serving biases are present, the $teams\ inefficiency$ $result$ implies that researchers may not want to write a joint paper even if it would be optimal to do so under a rational perspective. By anticipating the possibility that your coauthor will attribute success to his own talent rather than to your ability, you may decide to write a paper alone even if $\gamma > 1$. As long as researchers suffer from learning biases and take these biases into account, less joint papers will be written than in the absence of such biases. If researchers exhibit self-serving learning but do not realize neither their biases nor those of others (Assumption 2a), as many joint projects will be started in the self-serving and in the rational cases but a larger proportion will end after the first paper has been written.

5.1.3 A solution: find the optimal coauthor

In order to limit the negative effects of self-serving inference, researchers could select their coauthors according to their degree of self-serving biases. A possibility is to use gender, nationality or self-esteem as selection criteria to improve the choice of one's own coauthor. There is empirical evidence of significant cultural and gender differences

[34] This result can also be explained by assuming that agents have different priors about the probability of success on a given task (Van den Steen 2004).

in the magnitude of self-serving biases. Japanese have been found to be particularly responsive to negative signals about their ability whereas North Americans tend to discount such evidence (Kitayama et al. 1997, Heine, Kitayama, Lehman, 2001). Beyer (1990) found that self-serving evaluation biases are more representative of men than women. Another possibility for researchers is to select coauthors with low self-esteem. Psychologists have found that agents with low levels of self-esteem are equally likely to recall positive and negative self-relevant information (Kuiper and Derry 1982). Finally, establishing friendship among coauthors can reduce the negative effects of self-serving biases on cooperation. If workers learn as positively about others as they learn about themselves, self-serving biases will not undermine team formation. Brown (1986) found that friends and relatives are evaluated less negatively than an average person. However, the negative impact of self-serving biases on group formation does not disappear as long as learning biases are stronger in evaluating one's own ability than in assessing the ability of his coworker.

5.2 Applications to the organization

Cultural differences in the way people learn about themselves have been documented by psychologists. Japanese appear to be more self-critical than US and Canadian citizens (Kitayama et al. 1997, Heine et al. 1999, Heine, Kitayama, and Lehman 2001). In agreement with our model is the observation that the Japanese society characterized by self-criticism rather than self-serving attribution is associated with a corporate culture based on teamwork and cooperation (Abegglen 1958, Haitani 1990, Koike 1988). Florida and Kenney (1991) document that Japanese firms select their US employees taking into account their ability for teamwork. However, as it is argued by Sedikides, Gaertner and Vevea (2005) Japanese people may be self-enhancing in certain contexts. Assuming that Japanese are more self-critical than North American in the context of teamwork is in agreement with the conclusion of the meta analysis completed by Sedikides, Gaertner and Vevea (2005) in which they emphasize that Japanese tend to self-enhance on attributes relevant to the ideal of collectivism.

As a consequence of self-serving biases, team members may be willing to hire a third agent that would be able to provide objective and informed assessments about workers' abilities. Such an individual would be asked to design team contracts based on objective beliefs about workers' abilities. According to our theory of team formation, the need for such a third party is likely to be more pervasive when team workers are highly self-serving. This may explain why self-managed teams are less often observed in US organizations than in Japanese organizations (Koike 1988, Haitani 1990).

6 Conclusion

This paper explored the conditions for team formation when workers have the possibility to learn about their abilities. In particular, we analyzed the case of agents learning positively about themselves. This process is what psychologists refer to as self-serving learning or biased self-attribution. Interestingly, we showed that even if we considered sharing rules that are contingent on workers' relative abilities, learning has a negative impact on team formation. The intuition is that learning biases generate differences in beliefs about workers' abilities that may render impossible the design of an allocation rule that leads both workers to feel better-off working as a team. We established the robustness of the $teams\ inefficiency\ result$ to sophisticated individuals. We confirmed the negative effects of self-serving biases on team formation in the case of coworkers willing to overcome their biases by communicating through a revelation game. In the analysis of team contracts we showed that fixed allocation rules may be preferred to contingent contracts based on coworkers' abilities. We interpreted this result as the existence of psychological limitations on contracting.

We applied our model to the academic profession and concluded that too little cooperation among researchers may occur as a result of self-serving biases. Our model was finally used to understand organizational differences in Japan and in the US. In particular, Japanese organizations are known to be characterized by the intensive use of teamwork. We argued that self-criticism of Japanese employees can account for the extensive use of teams.

Also, we want to emphasize that solutions to insufficient cooperation may be very different whether limited cooperation is explained by individuals' self-serving biases or by asymmetry of information among agents. For example, in the former case it may be optimal to hide information to individuals whereas in the latter case information should always be released. As a result, we argue that it is important to identify whether inefficiencies have a psychological or an informational origin in order to select the adequate policies to reduce these inefficiencies.

Finally, our work, by establishing the negative impact of learning biases on cooperation in the workplace, challenges the view of researchers emphasizing the beneficial role of positive self-image.[35] It appears important to us to stress how positive self-image, despite its possible positive effects on individuals' motivation or mental health, can harm the society by undermining cooperation. To capture the possible negative effects

[35] Many authors have focused on how biased self-perception can increase motivation and lead agents to work harder. At a theoretical level we can refer to the works of Bénabou and Tirole (2002) and Gervais and Goldstein (2006). At the empirical level, Felson (1984) found that a positive view of oneself was associated to working harder and longer on tasks.

of biased self-attribution on the society, one may want to extend our model to more complex teams. A possible direction of research is to analyze the impact of learning biases on the formation of networks. A related investigation would be to assess the impact of these biases on the optimal organizational structure of firms.

7 Appendix

Proof of Proposition 1. The proposition follows from the conditions for team formation at $t = 1$ under the relative ability allocation rule. By comparing expected payoffs associated to team projects and individual projects, it is easy to see that teams are formed at $t = 1$ after a team has [not] been formed at $t = 0$ if $\gamma \geq \phi$ $[\frac{1}{\phi}]$. As a result, at $t = 0$, teams are formed whenever $\gamma \geq 1$. This is equivalent to say that the ETO is achieved. ■

Proof of Proposition 2. We show here that it is impossible to design allocation rules at $t = 1$ that lead workers to form teams whenever $\gamma \geq \frac{1}{\phi}$ $[\phi]$ after a team has [not] been formed at $t = 0$.

$i)$ At $t = 1$, if only one coworker suffers from self-serving biases and the performances are (B, B), the most favorable allocation rule for workers' cooperation, derived using the same reasoning as in the main text, is $\eta = \frac{1}{2}$. In that case team formation arises for $\gamma\phi \geq \frac{\phi\gamma_{BG}+1}{2} > 1$ since $\phi\gamma_{BG} = \frac{2q_G}{q_B+q_G} > 1$ (if a team has been formed at $t = 0$) and $\frac{\gamma}{\phi} \geq \frac{1}{2} + \frac{q_G}{q_B+q_G} \equiv \Theta_b > 1$ (if a team has not been formed at $t = 0$) .

$ii)$ If Worker 1 [Worker 2] suffers from learning biases and the history of signals is (B, G) $[(G, B)]$, the most favorable conditions for workers' cooperation, derived using the same reasoning as before, are such that $\eta_1 = \frac{2\alpha+1}{3\alpha+3}$. In that case, teams are formed for $\gamma\phi \geq \frac{1}{2} + \frac{q_G}{q_B+q_G} (\equiv \Theta_b) > 1$ (if a team has been formed at $t = 0$) and $\frac{\gamma}{\phi} \geq \Theta_b > 1$ (if a team has not been formed at $t = 0$).

$iii)$ If both workers suffer from learning biases and the history of signals is (B, B), the most favorable conditions for workers' cooperation are such that teams are formed for: $\gamma\phi \geq \frac{2q_G}{q_B+q_G} (\equiv \Theta_{bg}) > 1$ (if a team has been formed at $t = 0$) and $\frac{\gamma}{\phi} \geq \Theta_{bg} > 1$ (if a team has not been formed at $t = 0$). As a result, even selecting the allocation rule that maximizes workers' cooperation, the conditions for team formation at $t = 1$ are more demanding in terms of synergies than an ETE since $\Theta_b > 1$ and $\Theta_{bg} > 1$. ■

Robustness of Proposition 2 to general synergy functions. We have assumed up to now that the team outcome was $f(a, X_{1,t}, X_{2,t}) = a(X_{1,t} + X_{2,t})$, $\forall a \in \Re^m, \forall t \in \{1;2\}$ for $m \geq 1$. However, the inefficiency result captured in Proposition 2 is valid for more general synergy functions f. We denote:

$$\hat{f} \equiv \hat{f}(a, X_{1,2}, X_{2,2}, X_{1,1}, X_{2,1}) \equiv E[f(a, X_{1,2}, X_{2,2}) \mid X_{1,1}, X_{2,1}] \text{ and } q_{X_i} \equiv E[q_i \mid X_{i,1}]$$

and we assume that $\hat{f}$ and q_{X_i} are differentiable with respect to $X_{i,t} \in \Re$, $\forall i \in \{1;2\}, \forall t \in \{1;2\}$. We consider the existence of expected positive synergies for working in a team, that is $\hat{f}(a, X_{1,t}, X_{2,t}) \geq q_{X_1} + q_{X_2}$.

Biased self-attribution undermines team formation if (R) holds. We denote η_1 the share of the group outcome given to the first worker that allows teams to be formed in the absence of learning biases. We denote $\tilde{\eta}_1$ the allocation rule that is most favorable to team formation when learning biases are present. Worker i's biased self-attribution implies that $X_{i,t}$ is perceived as higher than its true value, $\forall i \in \{1;2\}$. System (R) corresponds to the case in which an efficient team would be formed in the absence of biased self-attribution but is not formed when learning biases occur.

$$(R) \Leftrightarrow \begin{cases} \eta_1 \hat{f}(a, X_{1,2}, X_{2,2}, X_{1,1}, X_{2,1}) \geq q_{X_1} \\ (1-\eta_1) \hat{f}(a, X_{1,2}, X_{2,2}, X_{1,1}, X_{2,1}) \geq q_{X_2} \\ \exists i \in \{1,2\} \ \textit{such that} \ \tilde{\eta}_i \left[\hat{f} + \frac{\partial \hat{f}}{\partial X_{i,t}}\right] < q_{X_i} + \frac{\partial q_{X_i}}{\partial X_{i,t}} \end{cases}$$

When expected synergies are taken to be sufficiently close to 0, (R) is approximately equivalent to $\left(\hat{R}\right)$:

$$\left(\hat{R}\right) \Leftrightarrow \begin{cases} \eta_1 \hat{f}(a, X_{1,2}, X_{2,2}, X_{1,1}, X_{2,1}) = q_{X_1} \\ (1-\eta_1) \hat{f}(a, X_{1,2}, X_{2,2}, X_{1,1}, X_{2,1}) = q_{X_2} \\ \exists i \in \{1,2\} \ \textit{such that} \ \eta_i \frac{\partial \hat{f}}{\partial X_{i,1}} < \frac{\partial q_{X_i}}{\partial X_{i,t}} \end{cases}$$

The first two equations are satisfied as long as workers are paid according to their relative expected abilities at $t = 1$. A consequence of the convergence of expected synergies to 0 is the convergence of $\tilde{\eta}_1$ to η_1. The last condition in $\left(\hat{R}\right)$ is equivalent to:

$\exists i \in \{1;2\}$ *such that* $\frac{\partial \hat{f}(a,X_{1,t},X_{2,t})}{\partial X_{i,t}} < \frac{1}{\eta_i}\frac{\partial q_{X_i}}{\partial X_{i,t}}$. As a result, Proposition 2 holds, for an arbitrarily small level of synergies, as long as the expected synergy function is not too steep with respect to performances. ■

Proof of Corollary 1. It follows directly from Propositions 1 and 2 since the ETO is achieved in the absence of self-serving biases under a relative ability allocation rule . ■

Proof of Corollary 2. *i*) The absence of a third party. As a result of Proposition 2, there exist no budget balanced contracts that can implement the ETO. In our framework a contract is budget balanced if and only if it is renegotiation-proof (this is a direct consequence of Lemma 1 in Bartling and von Siemens (2004) for $\beta = 0$). Since we consider budget balanced contracts ($\eta_1 + \eta_2 = 1$), there exist no renegotiation-proof contracts implementing the ETO.

ii) The case of a third party. We consider now the case $\eta_1 + \eta_2 < 1$ where a third party is assumed to be able to observe without biases the outcome of the two other workers. The third agent is rewarded a proportion η_3 of the group outcome $(\eta_1 + \eta_2 + \eta_3 = 1)$ and is considered to have the possibility to penalize team breaks by imposing a penalty on the agent(s) splitting the team at $t = 1$. A proportion $(1 - x)$ $\left[\left(1 - x^{\emptyset}\right)\right]$ of the aggregate payoffs of the first and second workers is given to the third party if a team is maintained [broken] in the second period, where $\left(x, x^{\emptyset}\right) \in [0,1]^2$. As a result, a worker retains a proportion x $\left[x^{\emptyset}\right]$ of his payoffs when [not] breaking a team at $t = 1$.

We consider the case in which a team has been formed at $t = 0$. The same reasoning applies if a team has not been formed at $t = 0$. We assume that $(X_{1,1}, X_{2,1}) = (B, B)$ and that the two workers suffer from biased self-attribution so they perceive their bad performance as if it was a good performance. This occurs with probability $p^2 p_{BB}$. A necessary condition to obtain the ETO is such that forming a team is preferred to working alone in that case. The conditions for team formation are stated in (Q).

$$(Q) \Leftrightarrow \begin{cases} \eta_1 \gamma\phi\left(q_B + q_G\right)x \geq x^{\theta} q_G \\ \left(1-\eta_1\right)\gamma\phi\left(q_B + q_G\right)x \geq x^{\theta} q_G \end{cases}$$

A necessary condition to get an ETE is then $\frac{x^{\theta}}{x\eta_1} = \frac{q_B+q_G}{q_G} = \frac{x^{\theta}}{x(1-\eta_1)} \Rightarrow \eta = \frac{1}{2}$. From the first inequality in (Q) we get that $x^{\theta} < x$ is a necessary condition for an ETE. However, for this contract to be renegotiation-proof, we need that the third party is better-off when a team is actually formed at $t = 1$ (if it was not the case, there would be an incentive to write a new contract a $t = 1$). This condition is $(Q') \Leftrightarrow \left\{\gamma\phi\left(1-x\right)2q_B \geq \left(1-x^{\theta}\right)2q_B\right.$

$\Leftrightarrow \gamma\phi \geq \frac{1-x^{\theta}}{1-x} > 1$, for $x^{\theta} < x$. If this condition is not satisfied, there exists a range of synergies and learning by doing effects such that a new contract preferred by the three agents can be designed at $t = 1$. This contract is devised such that x^{θ} is taken sufficiently high for (Q) not to hold. And, x^{θ} is taken to be inferior to x such that there exists $1 \leq \gamma\phi < \frac{1-x^{\theta}}{1-x}$ for which (Q') does not hold. As a result for $\gamma\phi \in \left[1, \frac{1-x^{\theta}}{1-x}\right[$, the three agents will be willing to write this new contract and the third party will let workers undertake the individual projects.

We conclude that there exist no long term commitment contracts that can implement the ETO since there exists a range of synergies higher than one such that the introduction of a third party does not imply team formation at $t = 1$ when the possibility of renegotiation is considered. ■

Proof of Proposition 3. We denote $\sigma_1 \equiv (\sigma_{11}, \sigma_{12})$ and $\sigma_2 \equiv (\sigma_{21}, \sigma_{22})$, and we take p_G $[p_B]$ to be the expected probability given information at $t = 0$ that $X_{i,1} = G$ $[X_{i,1} = B], \forall i \in \{1; 2\}$.

i) First, we show that an ETE is only possible if it is a truthful telling equilibrium (TTE). A truthful telling PBE is such that in equilibrium workers reveal their observed signals: $a_i = \sigma_i$ so that beliefs in equilibrium are such that $P\left[\left(X_{1,0'}, X_{2,0'}\right) = \left(a_{12}, a_{21}\right)\right] = 1$. Assume the payoff at $t = 0'$ is $\left(X_{1,0'}, X_{2,0'}\right) = (B, B)$ and both agents suffer from self-serving learning (i.e. $\sigma_1 = (G, B)$ and $\sigma_2 = (B, G)$). The ETO is implemented if team formation is obtained for any $\gamma\phi \geq 1$ $\left[\frac{\gamma}{\phi} \geq 1\right]$. We argue that these conditions for team formation can be obtained only if workers' beliefs converge in the revelation game. As long as agents' beliefs diverge, Proposition 2 shows that an ETE is no attainable. The only way beliefs can converge in the case mentioned above ($\left(X_{1,0'}, X_{2,0'}\right) = (B, B)$; $\sigma_1 = (G, B)$; $\sigma_2 = (B, G)$) is when both workers tell the truth. In that case, both workers learn that they performed poorly in the first period. As a result, an ETE has to be truthful telling.

$ii)$ Second, we prove that a truthful telling PBE cannot implement the ETO. This is the case since efficient teams ($\gamma \geq 1$) may not be formed when a team has [not] been formed at $t = 0$ for $\gamma\phi < \frac{2q_G}{q_B + q_G}$ $\left[\frac{\gamma}{\phi} < \frac{2q_G}{q_B + q_G}\right]$. A TTE must be such that workers cannot be worse-off by playing $a_i \neq \sigma_i$ whether a team has been formed at $t = 0$ or not. These conditions generate a system of 8 inequations that lead to the following unique solution $\eta_{(ijkl)} = \bar{\eta}, \forall\, (i, j, k, l) \in S^2$. The lower bound for achieving team formation is then$\gamma\phi \geq \frac{2q_G}{q_B + q_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2q_G}{q_B + q_G}\right]$ and corresponds to the case $\bar{\eta} = \frac{1}{2}$. Since $\frac{2q_G}{q_B + q_G} > 1$, we get the impossibility result stated in Proposition 3. ■

Proof of Proposition 4. $i)$ To prove the first part of the proposition we can easily check that, given $\eta\left(a_1, a_2\right) = \bar{\eta}$, the following strategies form a pooling equilibrium of our game : play $a_1 = (G, B)$ and $a_2 = (B, G)$, where the beliefs in equilibrium and out of the equilibrium are assumed to be the same as the prior beliefs. Team formation at $t = 1$ is attained for $\gamma\phi \geq M \equiv Max\left\{\frac{\hat{q}_G}{\eta(q_B + \hat{q}_G)}; \frac{\hat{q}_G}{(1-\eta)(q_B + \hat{q}_G)}\right\}$ $\left[\frac{\gamma}{\phi} \geq M\right]$ when a team has [not] been formed at $t = 0$, where $\hat{q}_G = wq_G + (1 - w)\, q_B$ with $w = \frac{p_G}{p_G + pp_B}$. This pooling equilibrium, being uninformative, is equivalent to an equilibrium obtained in the absence of a revelation mechanism in which at $t = 0$ workers decide on an allocation rule η for sharing the group outcome at the end of the first and second periods ($e.g.\ \bar{\eta} = \frac{1}{2}$).

$ii)$ The second part is established in the proof of Proposition 3. ■

Proof of Corollary 3. From Proposition 4 we know that the lowest bound for which the absence of a revelation mechanism can lead to team formation at $t = 1$ is $\gamma\phi \geq \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$ in the case $\eta = \frac{1}{2}$. Then, for $\gamma\phi \geq \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} \geq \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$ when a team has (not) been formed at $t = 0$, there exists no PBE that is preferred to the equilibrium obtained in the absence of a revelation mechanism when $\eta = \frac{1}{2}$. However, for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} < \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$, the revelation mechanism allows workers to achieve a more cooperative equilibrium in which both workers are better-off. As explained in the main text, for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ $\left[\frac{\gamma}{\phi} < \frac{2\hat{q}_G}{q_B + \hat{q}_G}\right]$ and $\eta = \frac{1}{2}$ teams are formed when signals are symmetric (this occurs with probability $p_{BB} + p_{GG}$ in the truthful telling PBE) whereas team are formed only when signals are symmetric and no biases have occurred ($(1 - p)p_{BB} + p_{GG}$) in the equilibrium obtained in the absence of a revelation game. ■

Proof of Proposition 5. $i)$ Truthful telling PBE associated to contracts (C^2_{TTE}) and (C^3_{TTE}). A TTE is such that $a_i = \sigma_i, \forall i \in \{1; 2\}$ where the beliefs in equilibrium $P\left[(X_{1,1}, X_{2,1}) = (a_{21}, a_{12})\right] = 1$. The truthful telling PBE associated to contract (C^2_{TTE}) leads to team formation for:

$(\sigma_1, \sigma_2) \in S^2$ Â $\{(B, B, B, G); (G, B, B, B); (G, B, B, G)\}$. The contract is defined, in the case a team has been formed at $t = 0$, as follows.

$$\left\{\begin{array}{c}
\gamma\phi\eta_{GBGB}\left(q_B+q_G\right)\geq q_G,\ \gamma\phi\left(1-\eta_{GBGB}\right)\left(q_B+q_G\right)\geq q_B \\
\gamma\phi\eta_{GBBB}2q_B<q_B,\ \gamma\phi\eta_{GBBG}2q_B<q_B,\ \gamma\phi\eta_{BBBG}2q_B<q_B \\
\eta_{GBGG}=\eta_{GBGB},\ \gamma\phi\eta_{BGBG}\left(q_B+q_G\right)\geq q_B \\
\gamma\phi\left(1-\eta_{BGBG}\right)\left(q_B+q_G\right)\geq q_G \\
\eta_{BGBG}=\eta_{GGBG},\ \gamma\phi\eta_{GGGG}2q_G\geq q_G \\
\gamma\phi\left(1-\eta_{GGGG}\right)2q_G\geq q_G,\ \gamma\phi\eta_{BBBB}2q_B\geq q_B \\
\gamma\phi\left(1-\eta_{BBBB}\right)2q_B\geq q_B,\ \eta_{ijkl}=0,\forall\left(i,j,k,l\right)\notin S^{\theta}
\end{array}\right\}\Leftrightarrow\left(C_{TTE}^{2}\right)$$

Where:

$$S^{\theta}\equiv\left\{\begin{array}{c}
\left(G,B,G,B\right);\left(G,B,G,G\right);\left(B,G,B,G\right);\left(G,G,B,G\right);\left(G,G,G,G\right); \\
\left(B,B,B,B\right);\left(B,B,B,G\right);\left(G,B,B,G\right);\left(G,B,B,B\right)
\end{array}\right\}$$

Contract $\left(C_{TTE}^{2}\right)$ leads to team formation with probability $p_{GG}+p_{BG}+p_{GB}+p_{BB}\left(1-p\right)^2$.

We consider an alternative truthful telling PBE leading to team formation for $(\sigma_1,\sigma_2)\in S^2$ Â (G,B,B,G). For $\gamma\phi\geq\frac{q_G}{q_B+q_G}+\frac{1}{2}$, this equilibrium is attainable under the following contract:

$$\left\{\begin{array}{c}
\eta_{GBGB}\in\left[\frac{1}{\gamma\phi}-\frac{q_B}{\gamma\phi(q_B+q_G)},1-\frac{1}{2\gamma\phi}\right] \\
\eta_{GBGG}=\eta_{GBBB}=\eta_{GBGB} \\
\eta_{GBBG}\in\left[0,\frac{1}{2\gamma\phi}\right[\\
\eta_{BGBG}\in\left[\frac{1}{2\gamma\phi},1-\frac{q_G}{\gamma\phi(q_B+q_G)}\right] \\
\eta_{GGBG}=\eta_{BBBG}=\eta_{BGBG} \\
\eta_{GGGG}\in\left[\frac{1}{2\gamma\phi},1-\frac{1}{2\gamma\phi}\right],\ \eta_{BBBB}\in\left[\frac{1}{2\gamma\phi},1-\frac{1}{2\gamma\phi}\right] \\
\eta_{ijkl}=0,\forall\left(i,j,k,l\right)\notin S^{\theta}
\end{array}\right\}\Leftrightarrow\left(C_{TTE}^{3}\right)$$

This contract leads to team formation for $(\sigma_1,\sigma_2)\in S^2$ Â (G,B,B,G) as long as $\gamma\phi\geq\frac{q_G}{q_B+q_G}+\frac{1}{2}$.

ii) Contracts $\left(C_{TTE}^{2}\right)$ and $\left(C_{TTE}^{3}\right)$ are the best possible contingent contracts.

To end the proof of Proposition 5, we need to show that there does not exist a contingent contract improving (C^1_{TTE}), (C^2_{TTE}) and (C^3_{TTE}). We show below how contingent contracts cannot achieve team formation for all $\sigma = (\sigma_1, \sigma_2) \in Z$ when $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$ where: $Z \equiv \{(B, G, B, G)\,; (G, B, G, B)\,; (G, B, B, G)\}$. This allows us to show that no contracts can be preferred to (C^2_{TTE}) and (C^3_{TTE}). The issue is that, for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$, team formation is obtained for all $(\sigma_1, \sigma_2) \in Z$ only if the allocation rule obtained when $\sigma = (B, G, B, G)$ is different from the allocation rule following $\sigma = (G, B, G, B)$. We show this impasse below in the three possible cases a), b) and c).

a) Worker 1 plays a different action after $\sigma_1 = (B, G)$ and $\sigma_1 = (G, B)$ whereas Worker 2 plays the same action $(\hat{a})$. Given $(X_{1,1}, X_{2,1}) = \sigma_1 = (B, G)$, and ensuring team formation for all $\sigma \in Z_1 \subset Z$, where $Z_1 \equiv \{(G, B, B, G)\,; (B, G, B, G)\}$, we need $\eta_{BG\hat{a}} = \eta_{GB\hat{a}}$. However, for $\eta_{BG\hat{a}} = \eta_{GB\hat{a}}$ no teams can be formed for both $\sigma = (G, B, G, B)$ and $\sigma = (B, G, B, G)$ when $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$.

b) Worker 2 plays a different action after $\sigma_2 = (B, G)$ and $\sigma_2 = (G, B)$ whereas Worker 1 plays the same action. Using the same reasoning than above we can show that, for $\gamma\phi < \frac{2\hat{q}_G}{q_B + \hat{q}_G}$, team formation is not possible for all $\sigma \in \{(G, B, B, G)\,; (G, B, B, B)\}$ when team formation is ensured for $\sigma \in \{(G, B, G, B)\,; (B, G, B, G)\}$.

c) Worker 1 [2] plays a different action after $\sigma_1 = (B, G)$ $[\sigma_2 = (B, G)]$ and $\sigma_1 = (G, B)$ $[\sigma_2 = (G, B)]$. If we want teams to be formed for all $(\sigma_1, \sigma_2) \in Z_2$, where $Z_2 \equiv \{(G, B, B, G)\,; (B, B, B, G)\,; (G, B, B, B)\}$ then we need to impose $\eta_{BGBG} = \eta_{GBBG} = \eta_{GBGB}$. However, for $\eta_{BGBG} = \eta_{GBBG} = \eta_{GBGB}$ team formation is not possible for all $\sigma \in \{(G, B, G, B)\,; (B, G, B, G)\}$ when team formation is imposed for $\sigma = (G, B, B, G)$. If we impose team formation only for $\sigma \in \{(G, B, B, B)\,; (B, B, B, G)\}$, we can get team formation as well for $(X_{1,1}, X_{2,1}) \in \{(G, B)\,; (B, G)\}$ as long as $\gamma\phi \geq \frac{q_G}{q_B + q_G} + \frac{1}{2}$.

iii) Determining the Best contracts. It follows from *ii*) that for $\gamma\phi < \frac{q_G}{q_B + q_G} + \frac{1}{2}$, the

Best contract is the best among (C^2_{TTE}) and $(C^1_{TTE,\frac{1}{2}})$, and for $\gamma\phi \geq \frac{q_G}{q_B+q_G}+\frac{1}{2}$, the Best contract is the best among (C^3_{TTE}) and $(C^1_{TTE,\frac{1}{2}})$. For $\gamma\phi < \frac{q_G}{q_B+q_G}+\frac{1}{2}$, (C^2_{TTE}) is preferred to $(C^1_{TTE,\frac{1}{2}})$ if:

$$\left\{\begin{array}{l}\text{Probability of team formation at } t=1 \text{ under } (C^2_{TTE}) \geq \\ \text{Probability of team formation at } t=1 \text{ under } (C^1_{TTE,\frac{1}{2}})\end{array}\right\} \Leftrightarrow \left(p^2-2p\right)p_B+2p_G \geq 0$$

Similarly, for $\gamma\phi \geq \frac{q_G}{q_B+q_G}+\frac{1}{2}$, (C^3_{TTE}) dominates $(C^1_{TTE,\frac{1}{2}})$ if $-p^2p_B+2p_G \geq 0$. ■

Proof of Corollary 4. *i&ii)* The first and second parts of the proof follow from the definition of the contracts. For $\gamma\phi < \frac{2\hat{q}_G}{q_B+\hat{q}_G}$, the Best fixed allocation rule contract is $(C^1_{TTE,\frac{1}{2}})$. This contract is dominated by contingent contracts, respectively by (C^3_{TTE}) and (C^2_{TTE}), for $p \leq \sqrt{\frac{2p_G}{p_B}}$ and $p\,(2-p) \leq \frac{2p_G}{p_B}$.

iii) The third part of the corollary follows from the existence of an uninformative PBE based on fixed allocation rules leading to team formation whenever $\gamma\phi \geq \frac{2\hat{q}_G}{q_B+\hat{q}_G}$ (Proposition 4). ■

8 References

Abegglen, J. 1958. *The Japanese factory,* Cambridge, MA, MIT press.

Babcock, L., Loewenstein G., Issacharoff S., and C. Camerer. 1995. "Biased judgments of fairness in bargaining." *American Economic Review,* 85(5): 1337-1443.

Babcock, L., G. Loewenstein. 1997. "Explaining bargaining impasse: the role of self-serving biases." *The Journal of Economic Perspectives,* 11(1): 109-126.

Bartling B., F. von Siemens. 2004. "Efficiency in team production with inequity averse agents." University of Munich Working Paper.

Bénabou, R., J. Tirole. 2002. "Self-confidence and personal motivation." *Quarterly Journal of Economics,* 117(3): 871-915.

Beyer, S. 1990. "Gender differences in the accuracy of self-evaluations of performance." *Journal of Personality and Social Psychology,* LIX: 960-970.

Box, G., G. Tiao. 1973. *Bayesian inference in statistical analysis*, Reading MA: Addison Wesley.

Bradley, G. W. 1978. "Self-serving biases in the attribution process: a reexamination of the fact or fiction question." *Journal of Personal and Social Psychology,* 36: 56-71.

Brown, J. D. 1986. "Evaluations of self and others; self-enhancement biases in social judgements." *Social Cognition*, 4: 353-376.

Camerer, C. F., A. Vepsalainen. 1988. "The efficiency of cultural contracting." *Strategic Management Journal,* 9: 77-94.

Caruso, E. M., Epley, N., and M., Bazerman. 2005a. "Leader of the packed: unpacking, egocentrism, and the costs and benefits of perspective taking in groups." Harvard University Working Paper Harvard University 05-035.

Caruso, E. M., Epley, N., and M., Bazerman. 2005b. "The good, the bad, and the ugly of perspective taking in groups." Harvard University Working Paper 05-036.

Che, Y., S. Yoo, 2001. "Optimal incentives for teams." *American Economic Review,* 91: 525-541.

Cooper, A., Woo, C., W. Dunkelberg. 1988. "Entrepreneurs' perceived chances for success." *Journal of Business Venturing*, 3: 97-108.

Eaton, B. C., A. Hollis. 2003. "Private information and teamwork." University of Calgary Working Paper.

Englmaier, F., A. Wambach. 2006. "Optimal incentives contracts under inequity aversion." IZA Working Paper 1643.

Epstein, S. 1983. "The unconscious, the preconscious, and the self-concept." In J. Suls and A. Greenwald (Eds.), Psychological perspectives on the self, vol2: 220-247, Hillsdale, NJ: Erlbaum.

Farrell, J. S., S. Scotchmer. 1988. "Partnerships." *Quarterly Journal of Economics,* 103(2): 279-297.

Felson, R. B. 1984. "The effects of self-appraisals of ability on academic performance." *Journal of Personality and Social Psychology,* 47: 944-952.

Fiske, S. T., S. E. Taylor. 1991. *Social cognition*, McGraw-Hill, 2nd edition.

Florida, R., M. Kenney. 1991. "Transplanted organizations: the transfer of Japanese industrial organizations to the US." *American Sociological Review,* 56: 381-398.

Gilbert, D. J., Pinel, E. C, Wilson, T. D, Blumberg, S. J., and T. Wheatley. 1998. "Immune neglect: a source of durability bias in affective forecasting." *Journal of Personality and Social Psychology,* 75: 617-638.

Gervais, S., I. Goldstein. 2006. "The effects of biased self-perception in teams." The Fuqua School of Business Working Paper.

Greenwald, A. G. 1980. "The totalitarian ego: fabrication and revision of personal history." *American Psychologist,* 35: 603-618.

Griffin, D., A. Tversky. 1992 "The weighing of evidence and the determinants of overconfidence." *Cognitive Psychology,*24: 411-435.

Haitani, K. 1990. "The paradox of Japan's groupism: threat to future competitiveness?" *Asian Survey,* 30: 237-250.

Hamilton, B., Nickerson, J., and H. Owan. 2003. "Team incentives and worker heterogeneity: an empirical analysis of the impact of teams productivity and participation." *Journal of Political Economy,* 111(3): 465-497

Hart, O., Moore, J. 2006. "Contracts as reference points." Harvard University Working Paper.

Heine, S. J., Lehman, D. R., Markus, H. R., and S. Kitayama. 1999. "Is there a universal need for positive self-regard?" *Psychological Review,* 106(4): 766-794.

Heine, S. J., Kitayama, S., D. R. Lehman. 2001. "Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information." *Journal of Cross-Cultural Psychology,* 32(4): 434-443.

Holmstrom, B. 1982. "Moral hazard in teams." *Bell Journal of Economics,* 13: 324-340.

Ishida, J. 2005. "Incentives in academics: collaboration under weak complementarities." OSIPP Discussion Paper.

Itoh, H. 1991. "Incentives to help in multi-agent situations." *Econometrica,* 59(3): 611-636.

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

Kitayama, S., Markus, H. R., Matsumoto, H., Norasakkunkit V. 1997. "Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan." *Journal of Personality and Social Psychology,* 72: 1245-1267.

Koike, K. 1988. *Understanding industrial relations in modern Japan*, New York, St Martin's.

Korner, I. 1950. "Experimental Investigation of some aspects of the problem of repression: repressive forgetting." N_0.970. New York, NY: Bureau of Publications, Teachers' College, Columbia University.

Kuiper, N. A., P. A. Derry. 1982. "Depressed and nondepressed content self-reference in mild depression." *Journal of Personality,* 50: 67-79.

Kuiper, N. A., M. R. McDonald. 1982. "Self and other perception in mild depressives." *Social Cognition,* 1: 233-239.

Kuiper, N. A., Olinger, L. J., McDonald M. R., B. F. Shaw. 1985. "Self-schema processing of depressed and nondepressed content: the effects of vulnerability on depression." *Social Cognition;* 3: 77-93.

Lee, P. M. 1997. *Bayesian statistics: an introduction*, 2nd edition, London: Arnold.

Lewinsohn, P. M., Mischel, W., Chaplin, W., R. Barton. 1980. "Social competence and depression: the role of illusory self-perceptions." *Journal of Abnormal Psychology,* 89: 203-212.

Liebowitz, S., J. Palmer. 1983. "Assessing assessments of the relative quality of Economics departments." *Quarterly Review of Economics and Business,* 28(2): 88-113.

Long, J. S., R. McGinnis. 1982. "On adjusting productivity measures for multiple authorship." *Scientometrics,* 4(5): 379-387.

McDowell, J., J. Smith. 1992. "The effect of gender-sorting on propensity to coauthor: implications for academic promotion." *Economic Inquiry,* 30: 68-82.

Miller, D. T., M. Ross. 1975. "Self-serving biases in the attribution of causality: fact or fiction?" *Psychological Bulletin,* 82: 213-225.

Mischel, W., Ebbesen, E. B., A. R. Zeiss. 1976. "Determinants of selective memory about the self." *Journal of Consulting and Clinical Psychology*, XLIV: 92-103.

Nisbet, R. E., L. Ross. 1980. *Human inference: strategies and shortcomings of social judgments*, Englewood Cliffs, NJ: Prentice-Hall.

Pronin, E., Lin, D. Y., L. Ross. 2002. "The bias blind spot: perceptions of bias in self versus others." *Personality and Social Psychology Bulletin,* 28: 369-381.

Rabin, M. 2002. "Inference by believers in the law of small numbers", *Quarterly Journal of Economics,* 117(3): 775-816.

Rey Biel, P. 2005. "Inequity aversion and team incentives." University College London Working Paper.

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

Salanié, B. 1997. *The Economics of contracts: a primer*, Cambridge Massachusetts: MIT Press.

Santos-Pinto, L, J. Sobel. 2005. "A model of positive self-image in subjective assessments." *American Economic Review,* 95(5): 1386-1402.

Schinski, M., Kugler, A., W. Wick. 1998. "Perceptions of the academic Finance profession regarding publishing and the allocation of credit in coauthorship situations." *Financial Practice and Education,* 8(1): 60-68.

Sedikides, C., Gaertner, L., J. L. Vevea. 2005. "Pancultural self-enhancement reloaded: a meta-analytic reply to Heine." *Journal of Personality and Social Psychology,* 89: 539-551.

Silverman, I. 1964. "Self-esteem and differential responsiveness to success and failure." *Journal of Abnormal and Social Psychology,* 69: 115-119.

Svenson, O. 1981. "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica,* 47: 143-148.

Van den Steen, E. 2004. "Rational overoptimism (and Other Biases)." *American Economic Review,* 94: 1141-1151.

Zuckerman M. 1979. "Attribution of success and failure revisited or, : the motivational bias is alive and well in attribution theory." *Journal of Personality,* 47: 245-285.