Incentives and the Structure of Teams*

April Franco[†], Matt Mitchell[‡] and Galina Vereshchagina[§]

August 2009

Abstract

This paper studies the relationship between moral hazard and the matching structure of teams. We show that team incentive problems may, in the absence of complementarities or anti-complementarities in production technology, generate monotone matching predictions. We also derive sufficient conditions on the primitives of the model leading to the optimality of positive and negative matching of team members.

Keywords: moral hazard, teams, assortative matching

JEL Classification Numbers: D21, D82, L23

*We thank participants of the seminars at the University of Toronto, Arizona State University, Penn State and Queens University, as well as the participants of Midwest Theory Conference at Ohio State University, 2009 Druid Conference, and 2009 Atlanta Competitive Advantage Conference, 2009 SED Annual Meeting and 2009 UBC Summer Conference on Industrial Organization.

[†]Rotman School of Management, University of Toronto

[‡]Rotman School of Management, University of Toronto

[§]Arizona State University

1 Introduction

Team production and team performance-based compensation schemes have been adopted by many firms over the last 30 years (as documented by Hamilton et al. (2003)). Given that employees usually differ in their skills and productivity, it is natural to ask whether the high skilled workers should be teamed with high skilled or low skilled partners. It is well known that supermodularity of payoffs leads to positive matching, while submodularity of payoffs induces negative matching (e.g., Becker (1973)). One plausible force behind the complementarity in payoffs is complementarity in the underlying technology (e.g. Kremer (1993)). In this paper we describe an additional explanation: the inability of employers to condition compensation on individual employees' contributions. In particular, we show that even when the production technology is modular, so that under complete information the principal is indifferent between the possible matching structures, the interaction between moral hazard and the team structure generates non-trivial matching predictions.

Our results are derived in a standard model of moral hazard in teams, in which a profit maximizing principal (firm owner) decides how to organize agents (workers) into teams. The team produces stochastic output (success or failure), and the probability distribution over output outcomes depends on the efforts exerted by the team members. Cost of exerting effort varies across workers and determines a worker's type. The principal observes the types of the workers and the team output but not the actual exerted effort and offers compensation scheme contingent on the agents' types and the team's realized output. In order to make the results as sharp as possible, we focus on a stark model, in which the principal has to form two teams of two out of four workers, two with a low cost of effort, and two with a higher cost of effort.

Then we develop three sets of results concerning the optimal matching structure when effort is not observable. First, we show that if types are ordered by the level of input they provide, the ordering of rewards (paid when successful output is observed) among those types determines the optimal matching structure. In particular, if higher input types are rewarded more in the event of a success, then negative matching is optimal. If higher input types are rewarded less, then the teams should be matched positively.

Intuitively, the team production structure generates payments that can be broken into two categories. In one category, the agent may be paid for a success for which his effort was responsible. In the other category, he may be also paid as a result of the effort of his teammate. We call this second category an "accidental" payment. To minimize the costs associated with accidental payments, the firm owner should team the high input worker with the one whose rewards are low, so that frequently paid accidental compensations are small in magnitude. Correspondingly, if the high input worker is promised a relatively low reward, positive matching is optimal; if, on the contrary, high input workers are paid high rewards, it is better to form mixed teams.

Then we go on to develop results about the impact of the shape of the underlying technology (the function describing the relationship between the team members' efforts and the probability of successful output) on the optimal matching structure. As is standard in the literature, incentive compatibility stipulates a particular relationship between the effort workers give and the reward given to the worker in the event of a success. This relationship is determined only by the properties of exogenous production technology and is derived from the incentive compatibility constraints. We show that, when input levels are interior, positive matching is associated with cases where the reward required to implement a particular effort level is a convex function of effort, while negative matching becomes optimal if the reward function is sufficiently concave.

This result is related to the primary result regarding the relationship between rewards and efforts. Consider the case where one agent has lower marginal and total cost. It is natural (and, indeed, optimal), for the agent with lower cost to be made to give more effort. But since that agent has lower cost, a particular level of effort can be achieved with a smaller reward. The question is how much more effort should be chosen for the low cost agent, and how much higher a compensation level that will necessitate. When the required reward is a very convex function of effort, it is prohibitively expensive to make effort substantially higher for the low cost agent, and therefore the main difference between the agents is not effort levels, but the required cost to provide a given effort level. As a result, the higher input agent (the low cost type) is paid less in the event of a success, and positive matching is optimal. By contrast, when the reward required to implement a given level of effort is a concave function, the low cost agent can be asked to give a great deal higher effort. In fact, so much so that the low cost agent gets a greater reward in the event of a success. Our earlier result about input levels and rewards implies that this case is exactly the one where negative matching is optimal.

Our final set of results includes ones concerning the effects of extreme effort assignments on matching predictions. We show that matching can go from positive to negative as parameters move into the region where lower corners become relevant. When effort of the high cost type becomes sufficiently expensive, it is optimal to not elicit from him any unobservable effort, and the principal would allocate low cost types to different teams in order to mitigate team incentive problems. By contrast, matching can go from negative to positive as upper corners begin to bind. Intuitively, when effort of the low cost type is very cheap, he is asked to exert the maximal feasible effort level (provided that such upper bound exists), but, at the same time, his reward may be very low. In particular, it becomes arbitrary close to zero and falls below the reward of the high cost type when the low-cost effort is sufficiently small. When this happens, our primary set of results about efforts and rewards implies that positive matching becomes optimal.

Although our results are developed in a principal-agent model with risk neutrality and limited liability, we extend the model to show that the intuition we develop applies in related settings. We discuss how the results relate to a case with either a looser limited liability constraint or a case where agents are risk averse and there is no limited liability. We show the sense in which the force we describe is part of those situations, although we show that such an environment also introduces a variety of other effects that make it analytically intractable. We present numerical results to show that similar results to the limited liability case can be obtained.

2 Related Literature

Our work is closely related to two strands of literature. First, moral hazard has been extensively discussed in the literature as a natural feature of team production. A large number of studies have analyzed the implications of the free-riding incentives as well as the properties of the optimal compensation to the team members (e.g., Holmstrom (1982), McAfee and McMillan (1991)). To our knowledge, however, our paper is the first attempt to understand whether moral hazard in teams may have certain implications for the optimal team structure in the presence of worker heterogeneity.

Second, there is a large body of work, dating back to Becker (1973), on matching patterns among heterogeneous agents. Recently, a number of studies analyzed how matching predictions might be affected by various economic frictions. For instance, one line of research pioneered by Shimer and Smith (2000) focuses on the role of search frictions and argues that in search models positive assortative matching may fail even if the joint production function is supermodular. Another friction that is likely to weaken the effects of technological complementarities is described by Kaya (2008), who illustrates that if the types of the matching partners are not observable, some of the low-type agents cannot be deterred from mimicking the high-type agents and, therefore, positive assortative matching cannot be sustained in the equilibrium. Our paper studies a different friction (moral hazard) and obtains quite different results: in contrast to the findings discussed above, we show that, depending on the parameters of the model, moral hazard can either reduce or increase the degree of technological complementarity embodied in production technology.

The effects of moral hazard on matching entrepreneurs to projects have been investigated by Thiele and Wambach (1998) and Newman (2007). They study a problem of assigning risk averse entrepreneurs with heterogeneous wealth to the projects with different amount of risk. In the absence of frictions, wealthier (and hence less risk averse) entrepreneurs would undertake riskier projects. However, if the entrepreneur's effort is unobservable, the opposite matching pattern may arise if the utility is linear in effort. This is because richer agents have lower marginal utility of income, and should be offered higher compensation to induce a particular effort level. While our paper also emphasizes the role of moral hazard, our question, as well as modeling environment, is very different from theirs, and the mechanism outlined in these papers does not play any role in generating our results. The matching in their papers is inter-firm, whereas we are concerned with matching within a given production process.

A more closely related paper to ours is a recent work by Kaya and Vereshchagina (2009) which analyzes the effect of moral hazard on the formation of partnerships in an equilibrium matching environment. There exist certain similarities between the driving forces leading to matching predictions in these two environments. While in our model matching patterns arise due to the interaction between the frequency and the size of accidental compensations, in their model matching predictions appear due to the interaction between the frequency and the size of inefficient punishments. Interestingly, under some restrictions on parameter values, these two environments generate opposite matching predictions, suggesting that different matching structures may arise in different organizational environments (small partnerships versus teams within large firms).

Finally, Prat (2002) poses a question related to ours, of how to organize workers in teams, but studies the effects of a different friction. The paper introduces learning into a team production model and shows that, in spite of this modification, the matching patterns are still determined by the super- or sub- modularities in the production technology. In contrast, in our model all the matching predictions arise solely due to moral hazard friction and are not driven by the properties of the underlying technology.

3 Model

Consider a technology owned by a risk neutral principal that requires two agents to be operable. Output from the technology is stochastic and depends on two inputs from each agent θ_i. The first input is binary for each agent, and it is Leontief in the sense that output is zero if either agent does not provide the input. This input will always be assumed to be observable, and therefore any contract will always specify that it be provided; its role will largely be suppressed.[1] Denote the second input (termed effort) by x_i. Assuming the first input is provided by both agents, output is (by normalization) one with probability $g(x_1) + g(x_2)$, where $g(x)$ is increasing and satisfies $0 \leq g(x) < 1/2$. Otherwise output is zero. We will term output of one a "success." Each agent is risk neutral and has cost of effort $c(x, \theta)$, where θ is the agents type. We assume that $c(x, \theta)$ is increasing in x and θ (i.e. higher types are less efficient) and $c(0, \theta) = 0$ for all θ.

Our choice of functional form for the success probability is driven by two considerations. First, we want the choice to be such that, in the absence of information frictions, matching is irrelevant. That is satisfied because of the additive separability in the inputs x_1 and x_2, as we will show formally below. Second, we want the function to be such that there is always some chance of failure, and therefore the incentive problem will be unavoidable when inputs are unobserved. This motivates the admissible range for $g(x)$.

In order to address the question of optimal team structure, we will suppose that the principal operates two teams and is faced with four agents, two each of types $\theta = \theta_l$ and $\theta = \theta_h$. The principal then has to decide whether to match like types (*positive* assortative matching) or different types (*negative* assortative matching), as well as a compensation scheme that specifies wages to each team's members contingent on observable variables. Since the teams do not interact in any way, it is sufficient to condition payments only on the variables relevant to the agent's team.

We normalize the agents' outside opportunity to zero and assume that they have limited liability, in the sense that the wages paid cannot be lower than zero in any state.[2] Since agents are risk neutral, it is sufficient to set wages in the event

[1] The role of this input is purely to insure that the technology describes a fundamentally "team" production problem. Both agents are essential.

[2] It is well known that in the absence of limited liability, risk neutrality of the principal

of failure to zero, and pay w_θ only in the event of success, as long as such payment scheme satisfies individual rationality.[3] The principal collects all output, net of wage payments. Formally, the principal can choose either to match positively, in which case he maximizes

$$2g(x_h)(1-2w_h)+2g(x_l)(1-2w_l) \tag{1}$$

or negatively, in which case his objective is

$$2(g(x_h)+g(x_l))(1-w_h-w_l). \tag{2}$$

4 Benchmark: Complete Information

A principal who could observe the inputs x, could make payments conditional on effort provision. Therefore, due to additivity in the production technology, the principal would simply choose x^*_θ for each agent to maximize

$$g(x^*_\theta)-c(x^*_\theta,\theta),$$

and pay wage

$$w_\theta=\frac{c(x^*_\theta,\theta)}{g(x^*_\theta)+g(x^*_{\theta^-})},$$

where θ^- is the the type of the agent's teammate.[4] Four agents, then, would generate expected output of $2g(x^*_l)+2g(x^*_h)$ and be paid expected wages of $2c(x^*_h,\theta_h)+2c(x^*_l,\theta_l)$, regardless of the matching structure. Thus, under complete information, the model does not generate any matching predictions.

---

and the workers implies that there would be no inefficiency associated with moral hazard: the optimal contract would induce the same level of effort as under full information, and the principal would collect the full information surplus net of the reservation values of the workers.

[3]Under limited liability and $c(0,\theta)=0$, the individual rationality constraint is indeed satisfied for any such contract.

[4]Of course, under full information, the compensation scheme does not have to be contingent on output. Instead, a constant wage of size $c(x^*_\theta,\theta)$ can be paid to the worker of type θ if the requested effort x^*_θ is exerted.

5 Incomplete Information

In the rest of the paper we focus on the case where only success or failure is observable, but inputs are not. In this case his efforts x_θ and wages w_θ must satisfies the incentive constraint

$$x_\theta \in \arg\max_x g(x)w_\theta - c(x,\theta) \tag{3}$$

An important feature of our additive structure for the underlying technology is that this incentive constraint is valid regardless of the team structure. Of course, the agent's compensation depends on the type of his partner, since the agent also collects $g(x_{\theta^-})w_\theta$. However, that is not relevant to the choice of x, and hence is left out of the incentive constraint. In other words, the agent's partner plays no role in the provision of incentives. Below we discuss alternative formulations where this is not the case, and argue that the intuition we develop for this stark model will naturally carry over.

Observe also that limited liability and $c(0,\theta) = 0$ guarantee that the worker's value $\max_x \left(g(x) + g(x_{\theta^-})\right) w_\theta - c(x,\theta)$ cannot be negative, thereby implying that the participation constraints are never binding. Thus the optimal contract offered by the principal maximizes (1) when the workers are sorted positively (or (2) when they are sorted negatively) subject to the incentive constraints (3) for all types. The optimal matching structure is, obviously, the one that delivers the highest profit to the principal.

It is important to note that both incomplete information *and* the team structure are essential ingredients to getting the matching results we develop. For instance, if the agents were in parallel moral hazard problems, so that a noisy signal of each agent's effort were available (rather than the joint signal from the team), then it is immediate that the problems of each agent are completely separable, and thus the principal's total profit does not depend on the matching structure.

5.1 Payment vs. Inputs

We first establish the basic result relating matching patterns with the correlation between effort and reward specified by the optimal contract.

Proposition 1 *Positive (negative) matching can be the optimal choice of the principal only if for such matching structure high-input types receive low (high) compensation in the event of success.*

Proof. Suppose to the contrary that the principal chooses to sort workers positively and offers higher compensation to the types exerting more effort. We will show that the principal's profit would increase if, instead, the workers were re-matched negatively.

For brevity, denote $g(x_l)$ and $g(x_h)$ by g_l and g_h. The total surplus of the firm owner matching his teams positively is equal to

$$\Pi^+ = 2g_l(1 - 2w_l) + 2g_h(1 - 2w_h).$$

As was noted above, the same contract will still be incentive compatible if the teams are re-matched negatively (since, due to additivity, incentive constraints are affected only by workers' own types). By re-matching the workers negatively, and offering the same contract, the firm owner would obtain profit

$$\Pi^- = 2(g_l + g_h)(1 - w_l - w_h).$$

It is easy to verify that

$$\Pi^- - \Pi^+ = 2(g_l - g_h)(w_l - w_h),$$

which is positive if high-input workers receive high rewards. Hence, the principal would benefit from re-matching the workers negatively, implying that positive matching can be chosen only if high input types receive low rewards.

The opposite result is based on a symmetric argument. ■

This result relates an endogenous (but possibly empirically observable) variable, relative pay across workers of different types, to the matching structure. The intuition is straightforward. Agents make "earned" income $g(x_\theta)w_\theta$ and "accidental" income $g(x_{\theta^-})w_\theta$. To keep accidental income at a minimum, agents with high inputs (i.e. high x, and therefore high $g(x)$) should be matched with agents with low wages. This insures that the relatively likely accidental payments are kept as small as possible. Note that the result does not require any special assumptions about the shape of $g(x)$ and $c(x, \theta)$.

Notice that the statement of Proposition 1 is conditional on the properties of the variables that are endogenously determined within our model, and it cannot be immediately stated how the efforts and rewards should be related to each other in the optimal incentive compatible contract. The reason is that, on one hand, the workers with the higher cost of effort should obtain a higher compensation for every unit of effort they provide, but, on the other hand, these workers would also be requested to exert lower effort level. These two effects act in the opposite directions and make the relationship between efforts and rewards ambiguous. In the following section we develop an example illustrating that, depending on the properties of $g(x)$, either effect may dominate, implying that moral hazard may potentially generate either positive or negative matching.

5.2 The shape of g and matching predictions: an example

To illustrate how the shape of $g(x)$ may be related to the optimal matching structure, consider the following example. Suppose that $c(x, \theta) = \theta x$ with $\theta \in \{\theta_l, \theta_h\}$ and $\theta_l < \theta_h$. Let $g(x)$ be piecewise linear, with slope a for low x, slope $b < a/2$ for intermediate values of x and slope 0 for large x. Suppose also that $a/2 > \theta_h > b$ and $b/2 > \theta_l > 0$. An example of such $g(x)$, $\theta_l x$ and $\theta_h x$ is illustrated on Figure 1. If θ_l is sufficiently close to 0, it is optimal to set x_l at the higher kink point and x_h at the lower kink point. The corresponding payments should be $w_h = \theta_h/a$ and $w_l = \theta_l/b$.[5]

To see why this shape of $g(x)$ may endogenously lead to either matching structure, consider adjusting x_h, while holding x_l, $g(x_l)$, and b constant. The left plot of Figure 1 gives an example of function $g(x)$, for which negative matching is optimal: when x_h is sufficiently low and, correspondingly, a is sufficiently high, the agent who contributes less receives smaller payment, since $w_h = \theta_h/a$ becomes smaller than $w_l = \theta_l/b$. If, on the other hand, x_h is sufficiently high so that a

[5]Note that $a/2 > \theta_h > b$ and $b/2 > \theta_l > 0$ imply that the principal's expected payoff is positive for either matching structure. This guarantees that the firm owner would not shut any of the teams down. It is still possible, however, that shutting one of the workers down would generate higher payoffs. Additional conditions may be imposed on the parameters to eliminate such possibility.

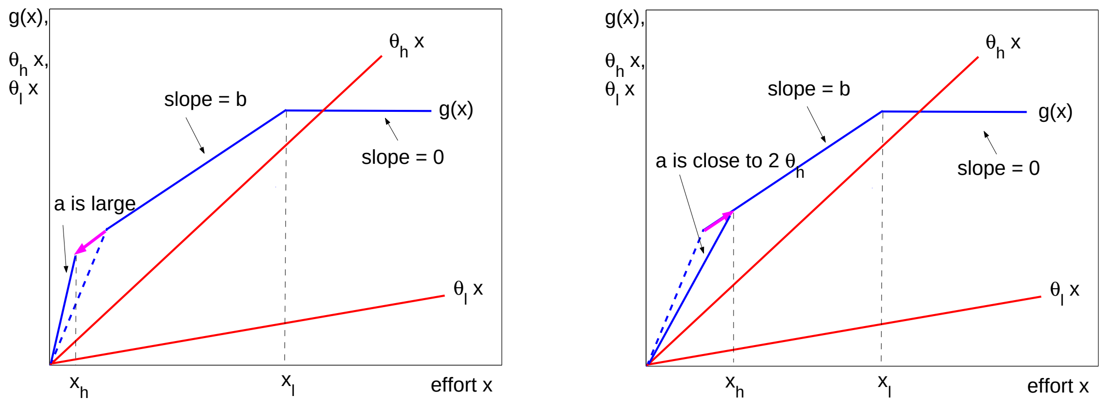

Figure 1: An example with piecewise linear $g(x)$. Parameters a and x_h may be adjusted to achieve either negative matching (on the left) or positive matching (on the right)

is arbitrary close to $2\theta_h$ (as illustrated on the right plot), the payment to the high-cost agent becomes bigger than the payment to the lower-cost agent, and positive matching becomes optimal.

5.3 The shape of g and matching predictions: interior effort assignment

Suppose that $g(x)$ is concave and three times continuously differentiable, and, as in the previous section, the cost function for type θ is $c(x, \theta) = \theta x$. In this case the incentive constraint for any interior effort level simplifies to

$$g'(x_\theta) w_\theta = \theta. \tag{4}$$

We can rewrite this as

$$w_\theta = \theta / g'(x),$$

or, letting $r(x) = 1/g'(x)$,

$$w_\theta = \theta r(x).$$

We interpret $r(x)$ as the shape of the reward as a function of effort.

Note that for $\theta_h > \theta_l$, any optimal contract must have $x_h < x_l$; otherwise the agents' inputs should be reversed.[6] To see how the shape of $r(x)$ is related to the matching structure, suppose there is negative matching. Our goal is to understand whether in this case the rewards specified by the optimal contract satisfy Proposition 1.

We can use the first order condition to write the principal's payoff just as a function of the inputs:

$$\max_{x_l, x_h} (g(x_l) + g(x_h))(1 - \theta_l r(x_l) - \theta_h r(x_h)) \tag{5}$$

The first order conditions are

$$\begin{aligned} g'(x_l)(1 - \theta_l r(x_l) - \theta_h r(x_h)) &= \theta_l r'(x_l)(g(x_l) + g(x_h)) \\ g'(x_h)(1 - \theta_l r(x_l) - \theta_h r(x_h)) &= \theta_h r'(x_h)(g(x_l) + g(x_h)) \end{aligned}$$

So we have that

$$\frac{g'(x_l)}{g'(x_h)} = \frac{\theta_l r'(x_l)}{\theta_h r'(x_h)}$$

or

$$\frac{\theta_h r(x_h)}{\theta_l r(x_l)} = \frac{r'(x_l)}{r'(x_h)}.$$

Since this is a negative match, the left hand side must be less than one, as it is the ratio of the pay of the low input type to the pay of the high input type. Thus negative matching is (strictly) optimal only if $r'(x_l) < r'(x_h)$, for which $r(x)$ must be concave. If, on the other hand, $r(x)$ is a convex function, then negative matching cannot be optimal as long as effort choices in heterogeneous teams are interior. Put in terms of $g(x)$, we have the following result:

Proposition 2 *Suppose that* $-\frac{g''(x)}{(g'(x))^2}$ *is (weakly) increasing and that in the negatively matched teams the effort assignment is interior. Then it is optimal to sort the teams positively.*

[6] If $x_h > x_l$ then by reversing inputs (while keeping the same matching structure) and adjusting the rewards to satisfy the agents' incentive constraints, the principal will be able to increase the expected profit. This can be easily verified by observing that concavity of $g(x)$ implies that $g'(x_h) < g'(x_l)$ and, therefore, in case of negative matching, $(g(x_h) + g(x_l))\left(\frac{\theta_h}{g'(x_h)} + \frac{\theta_l}{g'(x_l)}\right) > (g(x_l) + g(x_h))\left(\frac{\theta_h}{g'(x_l)} + \frac{\theta_l}{g'(x_h)}\right)$, and, in case of positive matching, $2g(x_h)\frac{\theta_h}{g'(x_h)} + 2g'(x_l)\frac{\theta_l}{g'(x_l)} > 2g(x_l)\frac{\theta_h}{g'(x_l)} + 2g'(x_h)\frac{\theta_l}{g'(x_h)}$.

A variety of commonly used examples meet this sufficient condition, including $g(x) = A(1 - e^{-Bx})$, $g(x) = A - B/x$, $g(x) = x - Ax^2$, and $g(x) = A + \ln x$.[7]

We can also prove a partial converse, relating concave $r(x)$ to negative matching.

Proposition 3 *Suppose that* $-\frac{g(x)g''(x)}{(g'(x))^2}$ *is (weakly) decreasing and for the positively matched team the optimal effort levels are interior. Then it is optimal to sort the teams negatively.*

Proof. For a team of two identical agents of type θ, the first order condition for the principal's problem is

$$g'(x)\left(1 - \frac{2\theta}{g'(x)}\right) + g(x) \cdot \frac{2\theta g''(x)}{(g'(x))^2} = 0. \tag{6}$$

It can be simplified to

$$\frac{g'(x)}{\theta} = 2\left(1 - \frac{g(x)g''(x)}{(g'(x))^2}\right). \tag{7}$$

or

$$\frac{1}{w_\theta} = 2\left(1 - \frac{g(x_\theta)g''(x_\theta)}{(g'(x_\theta))^2}\right)$$

If $-\frac{g(x)g''(x)}{(g'(x))^2}$ is decreasing the above equality implies that higher inputs receive higher rewards, which, by Proposition 1, is a contradiction to the asserted positive matching. ■

Proposition 3 implies that $r(x)$ must be sufficiently concave in order to get negative matching.[8] However, note that this does not mean that $g(x)$ must be terribly unusual; for instance, $g(x) = x^\alpha$ meets the requirement for all α strictly between zero and one.

[7] Another way to think about the condition is in terms of coefficients of absolute risk aversion and absolute prudence, taking $g(x)$ to be analogous to the utility function. Then the sufficient condition for negative matching is that coefficient of absolute prudence is larger than twice the coefficient of absolute risk aversion. Related conditions arise in other matching settings, such as Newman (2007) and Theile and Wambach (1999), where risk averse agents with unobservable effort are being sorted across projects with different level of risk.

[8] In the utility function language, the coefficient of absolute prudence, plus the ratio of $g'(x)$ to $g(x)$, must be smaller than twice the coefficient of absolute risk aversion.

Intuitively, the convexity/concavity of the reward function is important for matching predictions because it determines whether the differences in compensations are mostly driven by the differences in costs or by the differences in effort levels. As we emphasized earlier, the relationship between the agent's cost of effort and his compensation is generally ambiguous: an agent with the lower cost would be asked to exert more effort, but is compensated less per unit of effort. The question, then, is how much greater effort should be chosen for the low cost agent, and how much more reward that higher effort will necessitate. When the required reward is a very convex function of effort, it is prohibitively expensive to make effort substantially higher for the low cost agent, and therefore the main difference between the agents' rewards comes not from the exerted effort levels, but from the costs per unit of effort. As a result, the higher input agent (the low cost type) is paid less in the event of a success, and positive matching is optimal. By contrast, when the reward required to implement a given level of effort is a concave function, the low cost agent can be asked to give significantly higher effort relative to the high cost agent, eventually implying that the low cost agent gets a greater reward in the event of a success. The first set of results implies that this case is exactly the one where negative matching is optimal.

It is also instructive to point out that Propositions 2 and 3 establish sufficient conditions for super- and sub- modularity of the principal's profit function (under the assumption that solution is interior). This means that the mechanism described in our model illustrates how complementarities or anti-complementaries in the payoff function can endogenously arise due to unobservability of effort contributions from the individual team members. In addition, since our results hold for any θ_l and θ_h, where $\theta_h > \theta_l$, they can be immediately extended to the model with more than two types, as long as the technology is such that only the interior effort choices are made.

Although we focused on teams of two, these results for interior solutions can also be directly extended for larger teams. When solutions are interior, the matching structure can be described by the shape of the $g(x)$ function using exactly the conditions described in Propositions 2 and 3. The team size, then, can only impact the matching structure through the impact it has on choosing corners. As teams grow, the frequency of accidental compensations rises. Thus, in order to reduce the size of total accidental payments, there will be a tendency to move toward corner solutions, in which some of the team members provide no unobservable effort and receive no rewards in case of success. The following section analyzes the role of such corner solutions and argues that they may lead to switches from positive to negative matching.

5.4 The impact of the corner solutions on the matching structure

All the results in the previous section are formulated for interior solutions only. For some parameter values, however, the principal would choose extreme effort levels. Here we discuss how the matching structure may change if the corners arise.

Suppose that effort levels can be chosen from the interval $[\underline{x}, \overline{x}]$. Without loss of generality, we can normalize $\underline{x} = 0$. The upper bound $\overline{x}$ may appear due to technological restrictions or be stipulated by the restriction $g(x) < 1/2$. It can be easily seen that if $\lim_{x \to \underline{x}} g'(x)$ is finite and $\lim_{x \to \overline{x}} g'(x) > 0$ then, independently of the matching structure, both types of agents would be asked to exert zero unobservable effort if θ_l is sufficiently large, and maximum effort if θ_h is sufficiently low.[9] In either case the model would obviously have no matching predictions.

If, however, only one of the marginal costs is extreme, only one type of agent would be assigned an extreme effort, and the appearance of such corners may affect the matching structure chosen by the principal. In what follows we argue that particular corners tend to favor particular matching structures, namely that setting the effort of one of the agents to minimum creates incentives for the principal to form mixed teams, while setting the effort of one of the agents to maximum may induce the principal to rematch the mixed teams positively.

Intuitively, if the effort of the high-cost agent is very expensive, he would not be asked to exert any effort for any matching pattern, and only the low-cost types would be working. In this case it is optimal to assign the low-cost agents

[9] Note that corner solutions here need not be interpreted as agents not working. It is simply that those agents don't contribute unobserved effort and receive bonus payments. Agents with $x = 0$ could still be providing observable effort essential to the operation of the technology, and be compensated in accordance with their cost of effort. Here all those efforts and payments are fixed at zero, but none of the results require that those payments do not exist.

to separate teams in order to mitigate the team moral hazard problem. Thus the corner solutions where the high-cost types exert minimal effort may reverse the optimal matching structure from positive to negative. Alternatively, if the marginal cost of effort for low cost type is very low, the agent would be asked to exert maximal effort. Any further decrease in the marginal cost would not lead to higher exerted effort and, hence, would unambiguously lower the payment to this agent in case of success. Thus, if the parameters are such that in the interior solution higher effort is compensated with higher reward (i.e. negative matching is the optimal outcome), a significant decline in the marginal cost for one of the agents would increase his effort to the maximum $\overline{x}$ and drive his compensation down below his partner's. Consequently, by Proposition 1, switching the matching structure from negative to positive would reduce the total value of 'accidental' payoffs. We formally illustrate these intuitive arguments for particular forms of $g(x)$, exponential and power, for which, respectively, positive and negative matching is optimal in the interior solution.

5.4.1 The effect of the corner $x = 0$: exponential case

Suppose that $g(x) = (1 - e^{-x})/2$, with $x \geq 0$. Such $g(x)$ satisfies the condition in Proposition 2, and hence any interior case will always have the feature that matching is positive. We then only have to consider the outcome for this interior, positive matched case, as compared with the possible corner solutions for both matching structures.

For the interior case, the principal's surplus from a team of type θ is found by solving

$$\max_x 2g(x)(1 - 2\theta/g'(x))$$

The first order condition is[10]

$$e^{-x} = 4\theta e^{x} \tag{8}$$

which implies $e^{-x} = 2\sqrt{\theta}$ and the maximized payoff is

$$1 - 4\sqrt{\theta} + 4\theta. \tag{9}$$

[10] It is also shown in the Appendix that whenever positive matching is optimal, the second order condition for the positively matched teams is satisfied.

Note that such interior solution is possible only if $\theta < \frac{1}{4}$, for higher values of θ there exists no positive effort level which solves (8).

For comparison, consider a team with one team member giving zero effort. It is easy to show that the zero effort agent will always be the higher cost type, if the team is heterogenous. Therefore the principal must only choose an effort for the (weakly) lower type θ, in order to solve

$$\max_x g(x)(1 - \theta/g'(x))$$

In this case the first order condition is

$$e^{-x} = 2\theta e^{x}$$

so we have that $e^{-x} = \sqrt{2\theta}$ for the agent that works $x > 0$, and a maximized payoff of

$$1/2 - \sqrt{2\theta} + \theta \qquad (10)$$

In this case, the interior solution is possible only if $\theta < 1/2$.

A comparison of (9) and (10) implies that, for a team with two like types, the decision to choose $x = 0$ for one of them occurs at $\theta^* = \frac{3}{2} - \sqrt{2} < 1/4$, for small $\theta < \theta^*$, both work, for high $\theta \in [\theta^*, 1/2)$ only one works, and for $\theta \geq 1/2$ none of the agents works.

Obviously, if $1/2 \leq \theta_l < \theta_h$, no production takes place, and matching structure is irrelevant. It is also immediate that, if $\theta^* < \theta_l < 1/2$, so that any team would have at least one worker providing zero effort, it is optimal that the principal chooses negative matching, so that a low quality type provides effort in both teams. If, by contrast, $\theta_l < \theta_h < \theta^*$, positive matching has both agents on both teams working an interior amount. The alternative is negative matching with only the θ_l types working. The payoff in the positive case is

$$1 - 4\sqrt{\theta_l} + 4\theta_l + 1 - 4\sqrt{\theta_h} + 4\theta_h$$

whereas the negative case has payoff

$$2(1/2 - \sqrt{2\theta_l} + \theta_l).$$

It is easily verified that the positive matching payoff is larger for all θ_l and θ_h below θ^*.

The final case is the one where $\theta_l < \theta^* < \theta_h$. In this case, negative matching remains with the same payoff of

$$2(1/2 - \sqrt{2\theta_l} + \theta_l)$$

but positive matching gives a payoff of

$$1 - 4\sqrt{\theta_l} + 4\theta_l + 1/2 - \sqrt{2\theta_h} + \theta_h$$

since the team of two θ_h types has only one providing positive x. For fixed θ_l, as θ_h rises, the payoff to the positive matching case falls until the team is no longer productive at all (when $\theta = 1/2$), and the payoff to the negative matching outcome remains constant. For large h the negative case always dominates, and for θ_h near θ^* the positive case dominates. Therefore, for every θ_l, there is a cutoff, $\theta_h^*(\theta_l)$, such that negative matching is chosen if θ_h is greater than $\theta_h^*(\theta_l)$. It is easy to verify that $\theta_h^*(\theta_l)$ is decreasing in θ_l, with $\theta_h^*(\theta^*) = \theta^*$.

Intuitively, for fixed θ_l, negative matching is associated with a sufficiently high cost for higher cost workers. When both types are sufficiently good, all agents give effort and the form of $g(x)$ guarantees interior solutions. However, as the higher cost agent gets worse, the payoff from that team declines. Eventually the team with only θ_h types moves to a corner solution where only one agent provides effort. Eventually the payoff from that team doesn't even justify its existence, and the principal does better by switching to negative matching.

5.4.2 The effect of the corner $x = \overline{x}$: power case

Suppose that $g(x) = x^\alpha$, with $\alpha \in (0, 1/2)$ and feasible effort $x \in [0, \overline{x}]$. Then $-\frac{g(x)g''(x)}{(g'(x))^2} = \alpha(1-\alpha)$, which, by Proposition 3, implies that the optimal matching structure is negative. From the first order conditions, if interior effort levels are chosen, they must satisfy both

$$\alpha x_l^{\alpha-1} = \frac{\theta_l}{\alpha}\left[1+\left(\frac{\theta_l}{\theta_h}\right)^{\frac{\alpha}{1-2\alpha}}\right], \text{ and}$$
$$\alpha x_h^{\alpha-1} = \alpha x_l^{\alpha-1}\left(\frac{\theta_l}{\theta_h}\right)^{\frac{\alpha-1}{1-2\alpha}} = \frac{\theta_h}{\alpha}\left[1+\left(\frac{\theta_h}{\theta_l}\right)^{\frac{\alpha}{1-2\alpha}}\right]. \tag{11}$$

In the Appendix it is verified that the solution to (11) also satisfies the second order condition as long as $\alpha < 1/2$. It also follows from (11) that the rewards paid to the agents (found as $r_i = \theta_i/\alpha x_i^{1-\alpha}$) do not exceed 1/2, so the principal's profit is positive. Obviously, $\theta_l < \theta_h$ implies that $x_l > x_h$ as long as (11) holds. Also, since $\lim_{x\to 0} g'(x) = +\infty$, $x = 0$ is never chosen.

For a given θ_h, we find $\underline{\theta_l}(\theta_h)$ from

$$\alpha\overline{x}^{\alpha-1} = \frac{\underline{\theta_l}(\theta_h)}{\alpha}\left[1+\left(\frac{\underline{\theta_l}(\theta_h)}{\theta_h}\right)^{\frac{\alpha}{1-2\alpha}}\right].$$

As long as $\theta_h > 2\alpha^2\overline{x}^{\alpha-1}$, such $\underline{\theta_l}(\theta_h)$ is well defined and for a team $(\theta_h, \underline{\theta_l}(\theta_h))$ the effort assignment is $x_h < x_l = \overline{x}$ and (11) is satisfied. If θ_h is kept fixed and θ_l falls below $\underline{\theta_l}(\theta_h)$ then it is optimal to set $x_l = \overline{x}$ and $x_h = \min\{\overline{x}, \widehat{x_h}\}$, where $\widehat{x_h}$ solves

$$\alpha\widehat{x_h}^{\alpha-1}\left(1-\frac{\theta_l}{\alpha}\overline{x}^{1-\alpha}\right) - \frac{\theta_h}{\alpha} = \frac{\theta_h}{\alpha}(1-\alpha)\overline{x}^{\alpha}\widehat{x_h}^{-\alpha}. \tag{12}$$

The above equation is the principal's first order condition with respect to x_h when x_l is fixed at $\overline{x}$. In the Appendix we show that the solution $\widehat{x_h}$ to (12) exists, is unique, is strictly positive and that the second order condition is satisfied. It is also straightforward to see that as θ_l declines, $\widehat{x_h}$ rises and the reward paid to the high cost agent $r_h = \theta_h/\alpha\widehat{x_h}^{1-\alpha}$ increases (it converges to a positive value as $\theta_l \to 0$). Hence, there exists $\underline{\theta_l^*}(\theta_h) < \underline{\theta_l}(\theta_h)$ such that $r_l = \theta_l/\alpha\overline{x}^{1-\alpha} < r_h = \theta_h/\alpha\widehat{x_h}^{1-\alpha}$ for all $\theta_l < \underline{\theta_l^*}(\theta_h)$. In this case, since the agent exerting higher effort receives lower reward, Proposition 1 implies that the principal would obtain higher profit if he switches from negative to positive matching.[11] Note that, since Proposition

[11]Since $\widehat{x_h}$ solving (26) is decreasing in θ_l, it is possible that for $\theta_l < \underline{\theta_l^*}(\theta_h)$ the solution to (26) exceeds $\overline{x}$. This would be true if $\underline{\theta_l^*}(\theta_h) < \alpha\overline{x}^{\alpha-1} - \theta_h$. In this case, both agents are asked to exerts maximum effort, and there is no gain for the principal from switching to positive

1 provides necessary, but not sufficient, conditions, it may be optimal to switch to the positive matching structure even for some $\theta_l \in (\underline{\theta_l^*}(\theta_h), \underline{\theta_l}(\theta_h))$. In this case the gains from the switch would be driven by effort readjustment, not by reducing the accidental payments at given effort levels.

Finally, if $\theta_h \leq 2\alpha^2\overline{x}^{\alpha-1}$, then both agents in the mixed teams would be asked to exert maximal effort $\overline{x}$. If the agents were re-matched positively, the effort assignment for type θ would be given by $\min\left\{\overline{x}, \left(\frac{\alpha^2}{2\theta}\right)^{\frac{1}{1-\alpha}}\right\}$, which for $\theta_l < \theta_h \leq 2\alpha^2\overline{x}^{\alpha-1}$ also simplifies to $\overline{x}$. Hence, for these parameter values matching structure is irrelevant since both types exert $x = \overline{x}$ in any match.

We can summarize the above results as follows. If $\theta_h > \alpha^2\overline{x}^{\alpha-1}$ then there exists such $\theta_l^*(\theta_h)$ that negative matching is optimal for all $\theta_l \geq \theta_l^*(\theta_h)$, but it becomes optimal to rematch the teams positively if θ_l falls below $\theta_l^*(\theta_h)$. If θ_h is sufficiently low ($\theta_h \leq \alpha^2\overline{x}^{\alpha-1}$), both types exert maximum effort in either match, and the matching structure becomes irrelevant.

6 Relaxing Limited Liability

Our results were all generated in the context of a model where moral hazard frictions arise due to limited liability. Since agents are risk neutral, limited liability is essential to this friction. However our results are not driven by the restriction of non-negativity of rewards. Instead, we could require that $w_i \geq \underline{w}_i$, where $\underline{w}_i$ is negative (and potentially different across different types). Our result would still apply as long as, after the realization of 0 output, this constraint is binding for all the workers in all the matching structures. This would occur if $\underline{w}_i$ is not too small. When $\underline{w}_i$ becomes sufficiently low, the workers' individual rationality constraints would bind in all the matches for all the types. As a result, effort allocation would coincide with the first best, and matching would become irrelevant.[12]

Perhaps a better way to understand whether limited liability is the major

matching while keeping the efforts at the same levels. However, note that the principal's profit might still increase if, after rematching the agents positively, he asks some of them to exert less effort.

[12]The cases with intermediate values of $\underline{w}_i$, for which individual rationality constraint binds occasionally (for some types in some matches) are hard to characterize in general.

driving force of our results would be to consider a model in which the workers are risk averse, so that we can dispense with limited liability.[13] Suppose that each worker derives utility $u(w)$ from wage w, where $u(\cdot)$ is an increasing and concave function. The disutility of effort for type θ is given, as before, by $c(x, \theta)$. Under this modification, the decision problem of the firm owner becomes

$$\begin{aligned}
&\max_{x_i, x_j, w_i^1, w_i^0, w_j^1, w_j^0} [g(x_i) + g(x_j)]\,(1 - w_i^1 - w_j^1) - [1 - (g(x_i) + g(x_j))]\,(w_i^0 + w_j^0) \\
\text{s.t.} \quad & x_i = \arg\max_x \left\{ [g(x) + g(x_j)]\, u(w_i^1) + [1 - g(x) - g(x_j)]\, u(w_i^0) - c(x, \theta_i) \right\}, \quad i = l, h \\
& [g(x_i) + g(x_j)]\, u(w_i^1) + [1 - g(x_i) - g(x_j)]\, u(w_i^0) - c(x_i, \theta_i) = \underline{u}, \quad i = l, h \\
& w_i^1 \geq 0, \quad w_i^0 \geq 0, \quad i = l, h,
\end{aligned} \tag{13}$$

where w_i^1 and w_i^0 are the wage payments to type θ_i in case of success and failure respectively, and $\underline{u}$ is the workers' reservation value. Observe that limited liability constraint remains slack and the individual rationality constraint binds if $\lim_{w \to 0} u(w)$ is sufficiently small.

Unfortunately, in this setting it seems impossible to derive a result analogous to Proposition 1. In the benchmark model, the simple necessary conditions stipulating the relationship between inputs and rewards for a given matching structure were possible to obtain because in that environment the principal is able to adjust the matching structure while keeping efforts and compensations constant. Doing this has no effect on the expected output of the firm, but reduces the total compensation paid to the workers if the initial matching is suboptimal.

In contrast, in the model with binding individual rationality constraint and risk averse workers, the compensation scheme has to be adjusted if the principal rematches the teams while attempting to induce the same levels of effort. To see why this is the case, observe that, from the incentive constraint, the utility premium $u(w_i^1) - u(w_i^0)$ needed to induce effort x_i does not depend on the matching structure. The individual rationality constraint, however, stipulates that changing the type of the partner (and, hence, the probability of successful outcome)

[13] For concreteness, we formally study the case where compensation provided by the principal is the only source of the workers' consumption; nothing changes if the worker has outside income.

must affect $u(w_i^0)$. In particular, $u(w_i^0)$ (and hence w_i^0) must increase if, after rematching, the input of the worker's partner falls and must decline otherwise. In addition, due to concavity of $u(w)$, the spread $\Delta w_i = w_i^1 - w_i^0$ between the worker's compensations would also have to adjust in response to changes in w_i^0, because the spread $u(w_i^1) - u(w_i^0)$ between the utility levels has to remain constant to induce the original effort levels. More specifically, Δw_i must rise if w_i^0 rises and must fall otherwise. This implies that changing the matching structure would increase both compensations for one of the types and decrease them for the other. The effect on the total compensation is, therefore, ambiguous.

To see in more detail how these changes in the compensation structure impact the matching incentives of the principal, compare the principal's maximum profit for the positive matching structure

$$\Pi^+ = 2g(x_l) + 2g(x_h) - 2g(x_l) \cdot 2\Delta w_l - 2g(x_h) \cdot 2\Delta w_h - 2w_l^0 - 2w_h^0$$

with the profit that would be obtained if the workers were rematched negatively, without any change in effort levels

$$\Pi^- = 2g(x_l) + 2g(x_h) - 2(g(x_l) + g(x_h)) (\Delta\widehat{w}_l + \Delta\widehat{w}_h) - 2\widehat{w}_l^0 - 2\widehat{w}_h^0.$$

Here $\widehat{w}_i^1$ and $\widehat{w}_i^0$ are the wages of θ_i-type worker which induce effort x_i and satisfy the individual rationality constraint in the negative matching. Then

$$\begin{aligned}\Pi^+ - \Pi^- =&(g(x_h) - g(x_l))(\Delta w_l - \Delta w_h)+ \\ &\underbrace{\widehat{w}_l^0 - w_l^0}_{>0} + \underbrace{\widehat{w}_h^0 - w_h^0}_{<0} + (g(x_h) - g(x_l))(\underbrace{\Delta\widehat{w}_l - \Delta w_l}_{>0} + \underbrace{\Delta\widehat{w}_h - \Delta w_h}_{<0}).\end{aligned} \tag{14}$$

In the above expression, the first term corresponds to the interaction between inputs and rewards which drives the results of Proposition 1. Its presence indicates that the mechanism that leads to the matching predictions in the benchmark model is also present in this setting. The additional terms in the second row appear due to the case where individual rationality constraints bind and the concavity of the workers' utility function.[14] Their cumulative effect obviously de-

[14]If $x_l > x_h$ then the signs of the additional terms are exactly as marked in (14).

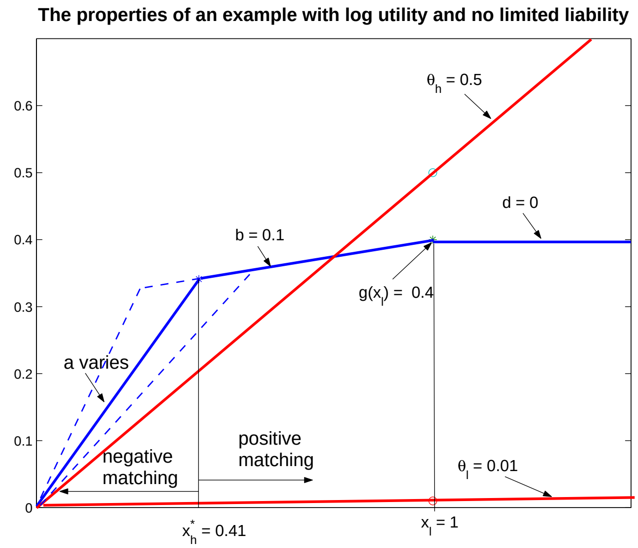

Figure 2: Matching predictions in the model with piecewise-linear $g(x)$ and logarithmic utility, $\underline{u} = \ln(0.05)$.

pends on the relative values of x_l and x_h, the curvature of $u(\cdot)$ and the properties of $g(x)$. That is why no sharp analytical predictions regarding the properties of the optimal matching structure can be derived in this environment.

We numerically solved an example with linear $c(x, \theta)$, piecewise-linear $g(x)$ (introduced in section 4.2) and logarithmic utility. We found that the model with risk aversion and binding individual rationality constraints generates predictions that are qualitatively similar to the ones obtained in our benchmark model: negative matching is optimal when x_h is sufficiently low, and positive matching is optimal when x_h is sufficiently high. Figure 2 illustrates an example for a particular set of parameters and reports the cutoff level x_h^* at which the principal is indifferent between the two matching structures.

7 Conclusion

We develop a model of team production where, in the absence of information frictions, there is no reason for any matching structure to prevail. Once we add a team moral hazard problem, typically there is a non-trivial matching decision for the principal. We formulate the solution to this decision in three ways. First we show that the decision is directly linked to the relationship between inputs across agents of different types, and the wage each receives. Second, we show that the structure is related to the shape of the relationship between an individual agents input and the team probability of success, in a way that is familiar from other matching structures that do not include a team component. Finally, for special cases, we show that the matching structure can depend on the way in which corner solutions arise. Corner solutions can switch the matching structure from positive to negative matching. This force seems to survive moving to a world with risk aversion by agents.

References

[1] G. Becker, A theory of marriage: Part I, Journal of Political Economy, 81 (1973) 813-846.

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

[3] B. Holmstrom, Moral hazard in teams, The Bell Journal of Economics, 13 (1982) 324-340.

[4] A. Kaya, Two-sided matching with private information, Working Paper, University of Iowa, 2008.

[5] A. Kaya, G. Vereshchagina, Moral hazard and equilibrium matchings in a market for partnerships, Mimeo, 2009.

[6] M. Kremer, The O-Ring theory of economic development, The Quarterly Journal of Economics, 108 (1993) 551-575.

[7] P. McAfee, J. McMillan. Optimal contracts for teams, International Economic Review, 32 (1991) 561-577.

[8] A. F. Newman, Risk-bearing and entrepreneurship, Journal of Economic Theory, 137 (2007) 11-26.

[9] A. Prat, Should a team be homogeneous?, European Economic Review, 46 (2002) 1187-1207.

[10] R. Shimer, L. Smith, Assortative matching and search, Econometrica, 68 (2003) 343-369.

[11] H. Thiele, A. Wambach, Wealth effects in the principal agent model, Journal of Economic Theory, 89 (1999) 247-260.

8 Appendix

8.1 Second order conditions for positively matched teams

If both workers in the team have marginal cost θ, then the principal's profit is given by

$$\Pi(x) = 2g(x)\left(1 - \frac{2\theta}{g'(x)}\right). \tag{15}$$

Correspondingly,

$$\Pi'(x) = 2g'(x)\left(1 - \frac{2\theta}{g'(x)}\right) + 4\theta g(x)\frac{g''(x)}{(g'(x))^2}, \tag{16}$$

and

$$\Pi''(x) = 2g''(x) - 2\theta\left(g(x)\frac{-g''(x)}{(g'(x))^2}\right)'_x. \tag{17}$$

Therefore, the second order condition is satisfied if $g(x)\frac{-g''(x)}{(g'(x))^2}$ is increasing (whenever the optimal matching structure is positive), but may be violated otherwise.

8.2 Second order conditions for negatively matched teams

The principal's profit in the team consisting of two different workers, with marginal costs θ_l and θ_h, is given by

$$\Pi(x_l, x_h) = (g(x_l) + g(x_h))\left(1 - \frac{\theta_l}{g'(x_l)} - \frac{\theta_h}{g'(x_h)}\right). \tag{18}$$

Thus

$$\frac{\partial \Pi(x_l, x_h)}{\partial x_l} = g'(x_l)\left(1 - \frac{\theta_l}{g'(x_l)} - \frac{\theta_h}{g'(x_h)}\right) + \theta_l \frac{g''(x_l)}{g'(x_l)^2}\left(g(x_l) + g(x_h)\right), \tag{19}$$

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l^2} = g''(x_l)\left(1 - \frac{\theta_h}{g'(x_h)}\right) - \theta_l \left(\frac{-g''(x_l)}{g'(x_l)^2}\left(g(x_l) + g(x_h)\right)\right)'_{x_l} \tag{20}$$

Note that, as for the positively matched teams, $\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l^2} < 0$ if positive matching is optimal but the second order condition may be violated if $g(x)\frac{-g''(x)}{(g'(x))^2}$ is decreasing (i.e., if the negative matching structure is optimal).
Since the problem is symmetric, the same is true for $\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_h^2}$.

The cross-derivative is

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l \partial x_h} = \theta_h g'(x_l)\frac{g''(x_h)}{(g'(x_h))^2} + \theta_l g'(x_h)\frac{g''(x_l)}{(g'(x_l))^2}, \tag{21}$$

which, by the first order conditions, can be simplified to

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l \partial x_h} = 2\theta_h g'(x_l)\frac{g''(x_h)}{(g'(x_h))^2} = 2\theta_l g'(x_h)\frac{g''(x_l)}{(g'(x_l))^2}. \tag{22}$$

Unfortunately, we cannot say anything conclusive regarding the sign of

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l^2}\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_h^2} - \left(\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l \partial x_h}\right)^2,$$

even if $\frac{-g''(x)}{(g'(x))^2}$ is weakly increasing (i.e. positive matching is optimal).

However, we can verify that the second order conditions hold for the example of the power function considered in Section 4.4.2. Plugging $g(x) = x^\alpha$ into (20) and (22), we find that

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l^2} = \alpha(\alpha - 1)x_l^{\alpha-2} < 0, \tag{23}$$

and

$$\frac{\partial^2 \Pi(x_l, x_h)}{\partial x_l \partial x_h} = 2(\alpha - 1)\theta_l x_l^\alpha x_h^{\alpha-1} = 2(\alpha - 1)\theta_h x_h^\alpha x_l^{\alpha-1}. \tag{24}$$

Correspondingly,

$$\frac{\partial^2\Pi(x_l,x_h)}{\partial x_l^2}\frac{\partial^2\Pi(x_l,x_h)}{\partial x_h^2}-\left(\frac{\partial^2\Pi(x_l,x_h)}{\partial x_l\partial x_h}\right)^2=$$
$$=\alpha^2(1-\alpha)^2\frac{x_l^{\alpha-1}x_h^{\alpha-1}}{x_l x_h}-4\theta_l\theta_h(\alpha-1)^2\frac{1}{x_l x_h}=$$
$$=\frac{(1-\alpha)^2}{x_l x_h}\theta_l\theta_h\left[\frac{1}{\alpha^2}\left(1+\left(\frac{\theta_l}{\theta_h}\right)^{\frac{\alpha}{1-2\alpha}}\right)\cdot\left(1+\left(\frac{\theta_h}{\theta_l}\right)^{\frac{\alpha}{1-2\alpha}}\right)-4\right], \tag{25}$$

which is definitely positive if $\alpha \in (0, 1/2]$.

8.3 Some properties of corner solution for power function

Consider the case when $\theta_h > 2\alpha^2\overline{x}^{\alpha-1}$ and θ_l is sufficiently small so that the principal chooses the corner solution $x_l = \overline{x}$. As mentioned in the main text of the paper, the optimal effort of the high cost agent in this case is $x_h = \min\{\overline{x}, x\}$, where x solves

$$\alpha x^{\alpha-1}\left(1-\frac{\theta_l}{\alpha}\overline{x}^{-\alpha}\right)-\frac{\theta_h}{\alpha}=\frac{\theta_h}{\alpha}(1-\alpha)\overline{x}^{\alpha}x^{1-\alpha} \tag{26}$$

Denote the left hand side and the right hand side of the above equation by $L(x)$ and $R(x)$ respectively. Obviously, both $L(x)$ and $R(x)$ are decreasing. Observe that $\alpha < 1/2$ implies that $\lim_{x\to 0} L(x)/R(x) = \infty$ and $\lim_{x\to\infty} L(x)/R(x) = 0$. Next, it can be verified that $\min_x L(x) - R(x) < 0$, implying that the above equation has at least one solution. To see that this solution is unique, it suffices to notice that $L'(x) > R'(x)$ implies that $L(x) < R(x)$. Hence, $L(x)$ and $R(x)$ have a unique intersection at which $L'(x) < R'(x)$, implying that there exists unique $x_h \in (0, +\infty)$ which maximizes the principal's expected payoff.