Incentives and the Structure of Teams*

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

May 2010

Abstract

This paper studies the relationship between moral hazard and the matching structure of teams. First, we show that team incentive problems may, on their own, generate monotone matching predictions in the absence of complementarities or anticomplementarities in production technology, and derive sufficient conditions on the primitives of the problem leading to the optimality of positive and negative matching of team members. Second, we analyze how the complementarity in the underlying technology affects the matching predictions which arise due to moral hazard. We find that, even when the underlying technology is strongly complementary, the incentive problem may lead to the formation of negatively sorted teams. Additionally, we show that the presence of moral hazard may break the usual link between complementarity and matching patterns: as the degree of complementarity increases, the optimal matching structure may switch from positive to negative, solely due to the need to provide incentives.

Keywords: moral hazard, teams, assortative matching

JEL Classification Numbers: D21, D82, L23

*We thank Hector Chade and 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 SED Annual Meeting and 2009 UBC Summer Conference on Industrial Organization. We benefitted greatly from the comments of an editor and two referees.

[†]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 (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. Since the supermodularity or submodularity of payoffs leads to positive or negative matching (Becker (1973)), one plausible force behind the matching structure is the shape of the underlying technology. For instance, in Kremer (1993), strong technological complementarities lead to positive matching. In this paper we highlight an additional force that may affect matching structure: 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 combination of moral hazard and the team structure generates non-trivial matching predictions. In addition, we show that when the production technology exhibits complementarities, positive assortative matching is not necessarily associated with higher degrees of complementarity. In particular, we document that, in contrast to the usual logic, the optimal matching structure can switch from positive to negative as the degree of complementarity increases, solely due to the presence of moral hazard.

The idea that there is interplay between moral hazard and team production dates back at least to the classic paper of Alchian and Demsetz (1972). They describe a team of two employees who work together to lift heavy objects; only their joint output is measurable. They focus on the question, further explored by Holmstrom (1982) in the context of partnerships, of how to reward teams for good outcomes in the face of free riding. We follow this line and explore a related question. Imagine a firm with workers of differing ability. How does the "moral hazard in teams" friction impact the optimal assortment of workers? For instance, in the Alchian and Demsetz heavy lifting example, should the firm combine its strongest employees on one team, or should they spread their strongest workers across teams?

Another well understood example in organizational economics that highlights the issues we consider in this paper is the interstate shipping industry (Nickerson and Silverman (2003)). Imagine a package that is shipped by two trucks as is seen in the case of less-than-truckload freight delivery. Both team members must exert unobserved care not to damage the contents. However, such damage, which is only observed at delivery, cannot be attributed to a specific driver. Should the firm tend to have packages handled on both ends by its best drivers or mix good drivers with lesser drivers on a given two-leg delivery? [1,2]

The shipping example underscores the impact of technological complementarities: it is pointless to waste a good quality driver on a package that has already been damaged or is likely to be. Therefore packages that ride with good drivers on the first leg should get good drivers on the second. This is the standard story of technological complementarities and matching in the literature.[3]

We highlight two new forces that arise as a result of the moral hazard problem, both stemming from the effect of moral hazard in teams on total compensation. First, in cases where employees are rewarded for successes, teammates' efforts increase the successes of the entire team, making everyone's expected compensation higher. The matching structure, therefore, matters to the firm because it impacts the total compensation of all team members. In section 3, in order to isolate this effect, we show that even when complementarities are absent from the underlying technology, this force alone can generate nontrivial matching predictions.

A second effect arises under complementarities. Consider an extreme case where good drivers always complete their portion of the trip successfully, so incentives (like success-contingent rewards) only need to be provided to the bad

[1] An alternative example is one of a moving company where there is a single truck, but different agents do the loading and the unloading.

[2] Other examples which fit the broad idea of moral hazard in teams and a choice of assortment patterns include sales teams and research teams within firms. Does a firm put together its best researchers on a project, or does it spread its best researchers across projects, when only the final output of the team is observable? These examples all fit the general spirit of the model we study.

Another transparent example, which has been the focus of empirical work on team production, is sports. Imagine a basketball team. Some of the team member's inputs are measurable, but a variety of inputs are not. The final output (winning) is the combination of efforts, including the unmeasurable ones, from a variety of team members. Matching structure is relevant: the coach must also decide whether to rest his best players simultaneously (so that the good players are always able to work together) or separately (so that the best players form teams with the inferior players).

[3] Formally, the probability of successful delivery is the product of the probabilities of successes at each leg, which makes production function supermodular.

drivers. There is a clear incentive benefit from negative matching: when a good driver is matched with a bad driver, the productivity of the latter increases (due to complementarity in production technology), which magnifies the impact of any bonus pay, making it easier to get incentives. Thus, compared to the case where bad drivers are matched with other bad drivers, their total expected compensation may actually go down if they are teamed up with good drivers, which might induce the firm owner to match the teams negatively. We study this force in the context of a technology with complementarities in section 4, and show that it can lead to negative matching even when technological complementarities are strong.

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 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 focus on the forces we emphasize, we focus on a stark model, in which the principal has to form two teams of two out of four workers, given that two of the workers have low cost of effort, and the effort of the other two is more costly.

We develop two sets of results concerning the optimal matching structure when efforts are not observable. The first set is for a technology that is modular (sections 3.2 and 3.3). We start by showing that if types are ordered by the level of input they provide, the ordering of rewards for successful output among those types determines the optimal matching structure. In particular, if higher input types are rewarded more, then negative matching is optimal. If higher input types are rewarded less, then the teams should be matched positively. The intuition is that the team production structure generates payments that can be broken into two categories. On the one hand, the agent may be paid for a success for which his effort was responsible. On the other hand, he may be also paid as a result of the effort of his teammate. We call this second kind of payment 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.

In addition, for the modular structure, we study the impact of the shape of the agent's cost functions 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 the agents' cost of effort functions and 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 concave.

Our second set of results focuses on the case when the underlying technology exhibits complementarities (sections 4.1 and 4.2). We show that in such cases negatively matched teams may still arise. This phenomenon is not limited to cases where complementarities are "small enough". We document this in two ways. First, we show that for any degree of complementarity we can find examples of types and cost functions that lead to negative matching. Second, for fixed types we show cases where the optimal matching structure is positive for little or no complementarity in the underlying technology, then switches to negative matching as the degree of technological complementarity increases, and then switches back to positive matching for extremely complementary technological specifications.

These results rely on a force not present with the modular technology. When efforts of the team partners are complementary, the amount of input from a team member's teammate impacts the agent's own marginal product of effort, thereby affecting the incentive compatible rewards that should be paid to him in case of success. In contrast, with the modular technology, the mapping from rewards to effort level chosen by an agent is independent of his teammate. Correspondingly, switching from positive to negative matching in the presence of complementarity increases the marginal rewards of the high input workers and lowers them for the low input workers. These changes have opposing effects on the total expected compensation. We derive the conditions on the cost function under which the latter effect is large relative to the former, so that the firm owner's profit is maximized when the workers are matched negatively. Interestingly, this occurs for the cost functions that favor positive matching in the modular case. The reason is that the change in the rewards due to complementarity is proportionate to the size of the rewards paid under positive matching. Thus, if under positive matching high input workers receive sufficiently low rewards, the increase in their reward after switching to negative matching is insignificant compared to the decline in the rewards of the low input workers. This is exactly the case when complementarity between the partner's efforts may result in negative matching (in contrast, recall that low rewards for high inputs are associated with positive matching in the modular case). We show that the negative matching can arise, and then disappear, as the degree of complementarity in the technology rises; in other words, matching structure is non-monotone in complementarities in technology.

Finally, in section 4.3 we use the results about negative matching with technological complementarities to show that the model is capable of explaining team production endogenously, while generating negative matching of the team members. Specifically, we show that it is possible to have the underlying technology specified in such a way that the technological complementarities are sufficient for team production to be superior to agents working without the benefit of a partner's input, but nonetheless the impact of the moral hazard means that negative teams dominate positive teams.

1.1 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 long literature, 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 received a lot of attention in the literature is the non-observability of the agents' types (as opposed to non-observability of their efforts, as is studied in our paper). For example, Meyer (1994) considers workers who start as "juniors" with unknown quality. Juniors are matched with experienced (senior) workers, and the firm decides whether to have juniors specialize to working for one senior, or be split across two senior's projects. This decision is motivated by learning: splitting junior agents across seniors makes is harder to learn the type of the junior agents, but makes it easier to ascertain the type of the senior agents. The optimal structure may be either specialization by juniors, or nonspecialization. Kaya (2008) studies a static matching market with two-sided private information and 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.

Anderson and Smith (2010) use a dynamic model with learning about unobserved types and show that positive matching might fail to be optimal even when the underlying technology exhibits complementarity. They argue that: "We believe that this is the first matching model since Becker (1973) where output is globally complementary and yet PAM fails." (Anderson and Smith (2010) p.1). The channel through which Anderson and Smith generate their results is the informational impact of matches, which drives the evolution of beliefs about agents' types. In our paper, the force that incomplete information brings is a static incentive effect. Our work is complementary to theirs. Our results, derived in a different sort of environment (static versus dynamic, moral hazard versus adverse selection), furthers the same line, that information frictions can drive matching structure. In addition, we show that moral hazard does not necessarily act against the existing technological complementarities. We illustrate that, instead, it may reinforce them, and provide conditions under which this would happen.

Matching predictions in the presence of two frictions, adverse selection and moral hazard, are analyzed in Jeon (1996). The paper studies agents of different generations who form teams. Agents have both unknown types and make unknown effort, but unlike our paper, most of the results are obtained for the case where wages are outcome-independent, and only dynamic reputational concerns drive agent's actions. Mixing old and young agents is beneficial because it makes it easier to ascertain the type of the young agent. Jeon (1996) also considers, for a given equilibrium level of wages, the way in which output can best be shared by agents in order to get incentives. With additive effort and types, Jeon finds that equal sharing is optimal. Unlike our model, the market determines total compensation and the agents decide how it should be divided, so the principal cannot choose compensation in order to elicit effort. This choice is the key in generating our results.

Thiele and Wambach (1998) and Newman (2007) investigate the effects of moral hazard on matching risk averse entrepreneurs with different wealth levels to 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.

An additional set of examples of team production includes partnerships, where a principal does not design a contract, but rather the team members design a sharing rule among them, with the same issues of free riding and team formation present. In a related paper, Kaya and Vereshchagina (2009) investigate the problem of partnerships in more detail and show that, in some cases, the matching predictions due to moral hazard arising in partnerships are the opposite to the ones we find in a principal – multiple agents problem.

2 Model

Consider a technology owned by a risk neutral principal that requires two agents to be operable.[4] Output from the technology is stochastic and depends on the inputs from each of the agents denoted by $x_i \in [0, \overline{x})$ (with $\overline{x}$ possibly infinite). Output is (by normalization) one with probability $f(x_1, x_2) < 1$, where f is symmetric and increasing in both arguments. With complementary probability output is zero. We will term output of one a "success." The restriction $f(x) < 1$ is imposed so that that there is always some chance of failure, and therefore the incentive problem is unavoidable when inputs are unobserved. This motivates the admissible range for $f(x)$. To make our points as transparently as possible we focus on the case where inputs are symmetric in f; this highlights differences in agents and not in their role in the technology.

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, convex, and three times differentiable in x. Further we assume that it satisfies $c(0, \theta) = 0$ for all θ. We normalize the agents' outside opportunity to zero[5] and assume that they have limited liability, in the sense that the wages paid by the principal cannot be below zero in any state.[6]

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 to choose a compensation scheme that specifies wages to

[4] We discuss later technologies where a team is endogenously beneficial.

[5] In general, the outside options of the workers do not have to be the same. For example, more skilled workers may also have higher outside options. It is important to point out that all our results remain unchanged if the outside options are type-specific but not binding, as in the model studied in the paper. If, however, the outside options for all types are restrictive and the individual rationality constraints bind in any matching structure, moral hazard would have no impact on matching predictions, since the total compensation would be determined by the value of the outside options and would be independent of the matching structure.

[6] It is well known that in the absence of limited liability, risk neutrality of the principal 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. In Franco, et al. (2009) we explore relaxing limited liability, and instead consider risk aversion in the agents' preferences.

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. Thus the effort assignment and compensation profiles in each team are chosen to maximize the principal's profit extracted from the team subject to incentive, individual rationality and limited liability constraints. The optimal matching structure is then chosen to maximize the principal's total expected profit. The principal's problem is formally set up in the following section.

3 Perfect Substitutes

In this section we focus on the special case where $f(x_1, x_2) = x_1 + x_2$ and $x \in [0, 1/2)$ so that the inputs of the two agents are perfect substitutes. This case is of particular interest because, as we show below, matching structure is irrelevant in the absence of information frictions, which allows us to isolate one effect of moral hazard on matching predictions arising in our main model.

3.1 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_i^* for the agent of type θ_i to maximize

$$x_i^* - c(x_i^*, \theta_i),$$

and pay wage

$$w_i = \frac{c(x_i^*, \theta_i)}{x_i^* + x_{i-}^*},$$

where x_{i-}^* is the the effort exerted by the agent's teammate, so that the worker's expected value is exactly equal to his outside opportunity.[7] Four agents, then, would generate expected output of $2x_l^* + 2x_h^*$ and be paid expected wages of $2c(x_h^*, \theta_h) + 2c(x_l^*, \theta_l)$, regardless of the matching structure. Thus, under complete

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

information, the model with perfect substitutes does not generate any matching predictions.

3.2 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 type θ_i's effort x_i and wages w_i must satisfy the incentive constraint

$$x_i \in \arg\max_x xw_i - c(x, \theta_i) \tag{1}$$

An important feature of the additive structure for the underlying production 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 – the agent also collects $x_{i^-} w_i$ – but 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 generalize the technology f to cases where this no longer applies.

The optimal contract offered by the principal to a team of workers (θ_1, θ_2), where $\theta_i \in \{\theta_1, \theta_2\}$ maximizes the expected profit subject to incentive, participation and limited liability constraints:

$$\begin{aligned} \Pi(\theta_1, \theta_2) = \max_{x_1, x_2, w_1, w_2} & (x_1 + x_2)(1 - w_1 - w_2) \\ \text{s.t. } & x_i \in \arg\max_x xw_i - c(x, \theta_i), \quad i = 1, 2 \\ & (x_1 + x_2)w_i - c(x_i, \theta_1) \geq 0, \quad i = 1, 2 \\ & w_i \geq 0, x_i \in [0, \overline{x}) \end{aligned} \tag{2}$$

Observe that, since agents are risk neutral, it is sufficient to set wages in the event of failure to zero, and pay compensation only in the event of success, as long as such payment scheme satisfies individual rationality. Thus the above decision problem assumes that agents receive wages only if the high output is realized. In addition, limited liability and $c(0, \theta) = 0$ guarantee that the worker's value $\max_x (x + x_{i^-}) w_i - c(x, \theta_i)$ cannot be negative, thereby implying that the individual rationality constraints never bind.

The optimal matching structure is, obviously, the one that delivers the highest profit to the principal. Thus the workers will be matched positively if $\Pi(\theta_1, \theta_1) + \Pi(\theta_2, \theta_2) \geq 2\Pi(\theta_1, \theta_2)$ and negatively otherwise.

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.

We first establish the following basic result relating matching patterns with the correlation between effort and reward specified by the optimal contract. For symmetric teams, the proposition is for teams which choose symmetric inputs for symmetric types; we provide sufficient conditions for symmetry below.[8]

Proposition 1 *(i) Positive matching with symmetric inputs within a team can be the optimal choice of the principal only if for such matching structure high-input types receive low compensation in the event of success.*

(ii) Negative matching can be the optimal choice of the principal only if for such matching structure high-input types receive high compensation in the event of success.

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 $x_i w_i$, which compensates their own effort, and "accidental" income $x_{i^-} w_i$, which compensates the effort of their partners. To keep accidental income at a minimum, agents with high inputs 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 $c(x, \theta)$.

The statement of Proposition 1 is conditional on the properties of the variables that are endogenously determined within our model. Notice also that it cannot be immediately stated how the efforts and rewards should be related to each other in

[8] The proofs of all results are contained in the appendix.

the optimal incentive compatible contract. The reason is that, on the 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 illustrate that, depending on the properties of $c(x, \theta)$, either of the two effects may dominate, implying that moral hazard may potentially generate either positive or negative matching for a modular technology.

3.3 The shape of $c(x, \theta)$ and matching

For the remainder of the paper we suppose that the cost function for type θ is $c(x, \theta) = \theta c(x)$. We will focus on the matching predictions for interior effort levels in this section, thus we begin by describing sufficient conditions for interiority and uniqueness, as well as symmetry of positively matched teams.[9]

Lemma 1 *For all* θ_1, θ_2*, there exists* $\varepsilon > 0$ *such that, if* $c'(0) < \varepsilon$ *and* $\lim_{x \to \bar{x}} c'(x) > 1/\min\{\theta_1, \theta_2\}$ *then*

(i) if $\theta_1 = \theta_2$*, then* $c'''(x) > 0$ *implies that solutions are unique, interior, and symmetric*

(ii) if $\theta_1 \neq \theta_2$*, then* $c''(0) = 0$ *implies that solutions are interior, and* $c'''(x) > 0$ *implies solutions are unique.*

The assumption on $c'(\bar{x})$ guarantees that employing an input at $\bar{x}$ would be unprofitable. Without $c'(0)$ sufficiently small, there is essentially a fixed cost of effort that can lead to corners at zero: inducing any positive effort requires a wage of at least $\theta c'(0)$. The assumption on c''' is the analog of the usual condition that guarantees concavity; as is customary in principal-agent problems, the first derivative enters the principal's problem through the first order condition of the agent, and two derivatives of that problem lead to concavity. The assumption of $c''(0) = 0$, which is required to guarantee interiority in negatively matched teams, is needed because the modular technology is not infinitely steep near zero. It can, however, be replaced by a sufficient condition on workers' types that would allow

[9] The role of corner solutions is briefly discussed in the end of this section and is studied in more details on Franco et al. (2009).

$c''(0)$ to remain positive; namely, for x_1 chosen by the worker of type $\theta_1 < \theta_2$, it is sufficient that

$$c''(0) < \frac{\theta_1}{\theta_2} c''(x_1)$$

Obviously, this condition holds for θ_1 sufficiently close to θ_2 when $c'''(x) > 0$[10], so interiority in negatively matched teams can be guaranteed without relying on $c''(0) = 0$ when the differences between the types are sufficiently small. No assumption on $c''(0)$ will be essential in the next section, where we study supermodular technologies with the usual Inada-type condition.

In the interior case the incentive constraint for any interior effort level simplifies to

$$w_i = \theta_i c'(x_i). \tag{3}$$

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

Recalling that the participation constraint never binds, the decision problem (2) of the principal modifies to

$$\Pi(\theta_1, \theta_2) = \max_{x_1, x_2} (x_1 + x_2)(1 - \theta_1 c'(x_1) - \theta_2 c'(x_2)) \tag{4}$$

To see how the shape of $c'(x)$ is related to the matching structure, consider an interior solution where there is negative matching, i.e. the planner operates two teams (θ_1, θ_2). Our goal is to understand whether in this case the rewards specified by the optimal contract satisfy Proposition 1. From first order conditions it is immediate that

$$\frac{\theta_2}{\theta_1} = \frac{c''(x_1)}{c''(x_2)}$$

and thus

$$\frac{\theta_2 c'(x_2)}{\theta_1 c'(x_1)} = \frac{c''(x_1)/c'(x_1)}{c''(x_2)/c'(x_2)} \tag{5}$$

Note that the left hand side of (5) is the ratio of the compensations paid to types θ_2 and θ_1 respectively. Thus, if $c''(x)/c'(x)$ is increasing in x then (5) implies that, for the interior solution, higher inputs receive lower compensation,

[10] Here x_1 solves the first order condition of the planner who makes only worker 1 exert effort, i.e. $1 - \theta_1 c'(x_1) = x_1 \theta_1 c''(x_1)$. Note that sufficiently small $c'(0)$, namely $c'(0) < 1/\theta_1$, guarantees that x_1 is bounded away from 0, which ensures that $c''(0) < \frac{\theta_1}{\theta_2} c''(x_1)$ indeed holds when θ_1 is close to θ_2.

in which case, according to Proposition 1, it would be optimal to rematch the workers positively. Thus we have the following result:

Proposition 2 *Suppose that* $\frac{c''(x)}{c'(x)}$ *is increasing for some* $x \in (x_L, x_H)$ *and that* θ_1 *and* θ_2 *are such that under negative matching optimal effort levels are chosen within* (x_L, x_H)*. Then it is optimal to sort the workers positively.*

An immediate implication of Proposition 2 is that if $\frac{c''(x)}{c'(x)}$ is increasing in the whole range of feasible x then positive matching is optimal for any combination of workers' types. An example of a function for which $\frac{c''(x)}{c'(x)}$ is increasing, and all the assumptions from Lemma 1 guaranteeing interiority hold, is $c(x) = A \arcsin(x/\bar{x})$ (for sufficiently small A). However, it is generally difficult to have $c''(x)/c'(x)$ increasing everywhere, since $c'(0)$ small makes the ratio high near zero. However, many functions have c''/c' diverging for x sufficiently close to $\overline{x}$, and for such functions x^2c meets all the assumptions of Lemma 1 at zero and, for types θ such that optimal x is in the increasing region of c''/c' (i.e. θ sufficiently small), must have positive matching. A variety of examples have c''/c' diverging, including $c(x) = -\ln(\bar{x} - x)$, $c(x) = e^{x/(\bar{x}-x)} - 1$, and $c(x) = \bar{x}^\alpha - (\bar{x} - x)^\alpha$ with $\alpha \in (0, 1)$. Note that by choosing, for instance, α one can make the range where c''/c' increases large, and therefore the range of θ for which positive matching is optimal large.

Regions where $c''(x)/c'(x)$ increases in x can be interpreted in terms of the shape of rewards. Interpreting $c'(x)$ as the reward as a function of x, this says that rewards are sufficiently convex in x (namely, log-convex): its derivative $c''(x)$ must be growing with x faster than $c'(x)$ does. We can follow a similar line to prove a partial converse, relating (sufficiently) decreasing $c''(x)/c'(x)$ to negative matching.

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

Proposition 3 implies that the reward $c'(x)$ can not be too convex in order to get negative matching. In particular, log concavity of $c'(x)$ is necessary but not sufficient. Rewards must be more-than-log concave. At the same time, this does not mean that $c(x)$ must be terribly unusual. For example, $c(x) = Ax^{\alpha}$ with $\alpha > 2$ satisfies this condition, and satisfies the assumptions in Lemma 1 so long as A and α are chosen so that $c'(\bar{x})$ is high enough.

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,[11] but is compensated less per unit of effort. The question, then, is how much greater a level of effort should be chosen for the low cost agent, and how much more cost 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 a great deal 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 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 $\theta_h > \theta_l$, 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

[11] Note that for $\theta_h > \theta_l$, any optimal contract must have $x_h \leq x_l$. 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 convexity of $c(x)$ implies that $c'(x_h) > c'(x_l)$ and, therefore, in case of negative matching, $(x_h+x_l)(\theta_h c'(x_h) + \theta_l c(x_l)) > (x_h+x_l)(\theta_h c'(x_l) + \theta_l c(x_h))$, and, in case of positive matching, $2x_h\theta_h c'(x_h) + 2x_l\theta_l c'(x_l) > 2x_h\theta_h c'(x_l) + 2x_l\theta_l c'(x_h)$.

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 $c(x)$ function using exactly the conditions described in Propositions 2 and 3.[12] 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.[13]

4 Supermodular f

A natural case to study in team production models is one in which the team members inputs are complements; this force can give rise to the formation of teams as an endogenous result, and not simply an assumption that the technology requires two agents. Given that we can see transparently in the modular technology case that moral hazard may affect matching predictions, in this section we study more general functions f, in order to understand how moral hazard interacts with technological complementarities in forming matching predictions. We will term our assumption of supermodular f as "technological supermodularity." We use this term to distinguish the assumption from the conclusion of positive or negative matching, which implies supermodularity or submodularity of the principal's payoff function.

[12]It is straightforward to verify that, under the positive matching structure, the interior solution for teams of more than two members is still symmetric. Namely, the planner maximizes $(x_1+...+x_N)(1-\theta_1 c'(x_1)-...-\theta_N c'(x_N))$. The first order condition with respect to x_i implies that $1-\theta_1 c'(x_1)-...-\theta_N c'(x_N)=(x_1+...+x_N)\theta_i c''(x_i)$. Since its left hand side does not depend on i, $\theta_1=..=\theta_N$ in conjunction with strict monotonicity of $c''(x)$ implies that $x_1=...=x_N$.

[13]Franco et. al. (2009) analyze in detail the impact of such corners on the matching structure and show that their appearance may lead to switches from homogeneous (positively matched) to mixed teams. Intuitively, this happens because if the principal chooses not to solicit any unobservable effort from some of the members of a team of low cost agents, he would rather prefer to assign these low cost agents to other teams where they could be used more productively, and instead replace them with high cost workers, who would not need to exert any unobservable effort anyway.

4.1 Technological Supermodularity and Negative Matching

Our first question is how the degree of technological complementarity in f impacts the effects of moral hazard on the matching structure. In the complete information case, supermodularity of f immediately implies positive matching is optimal. Therefore we focus our question by asking when negative matching can occur despite supermodularity of f.

A very first answer to the question can be taken from the prior section. We found that, in the case where the production function was modular, there were cost functions $c(x)$ such that it was *strictly* optimal to match negatively. Therefore, for a small amount of supermodularity in f, by continuity it is still the case that negative matching dominates positive matching. That is, there should be negative matching if supermodularity is not "too great."

The purpose of this section, however, is not to make that argument. Instead, we focus on a different channel, one that is not present in the additive case. The effect comes through the change in wages required to generate effort by an agent as his partner's type changes. With complementarity, when a worker's partner's input is increased, the worker's marginal product of effort goes up, and therefore effort is less expensive to elicit for the planner. In what follows we explore whether the effect of supermodularity in f on the compensation scheme may indeed give rise to negative matching. Throughout, as before, we focus on the case of interior solutions, and ask the same question: would switching from positive to negative matching, while keeping the inputs constant, make the principal better off? This is a sufficient condition for negative matching, but of course not necessary; it does, however, allow us to characterize regions where negative matching arises despite (possibly strong) technological complementarity.

In the rest of this section we assume that the production function f satisfies the following assumptions:

Assumption 1 *f is twice continuously differentiable with* $\lim_{x_1 \to 0} f_1(x_1, x_2) = +\infty$ *and* $\lim_{x_2 \to 0} f_2(x_1, x_2) = +\infty$ *for* $x_1, x_2 \in [0, \overline{x}]$.

Assumption 2 *For all* $x_1, x_2 \in [0, \overline{x}]$ *such that* $x_1 > x_2$ *at least one of the two conditions holds: either*

(i) $\frac{f_{12}(x_1,x_2)-f_{11}(x_1,x_2)}{f_1(x_1,x_2)^2} \geq \frac{f_{12}(x_1,x_2)-f_{22}(x_1,x_2)}{f_2(x_1,x_2)^2}$ *or*

(ii) $\frac{f_{12}(x_1,x_2)-f_{11}(x_1,x_2)\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)}}{f_1(x_1,x_2)} \geq \frac{f_{12}(x_1,x_2)-f_{22}(x_1,x_2)\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}}{f_2(x_1,x_2)}$.

The first assumption guarantees that the optimal effort choice is strictly positive. The second assumption, combined with $c'''(x) > 0$, is used to guarantee that if the planner sorts workers positively then the workers of equal types exert equal effort levels (i.e. the solution to the planner's decision problem is symmetric), which is later used to show that negative matching can arise under any degree of complementarity.

It is straightforward to verify that for the CES function $f = (0.5x_1^a + 0.5x_2^a)^{\frac{1}{a}}$ Assumptions 1 and 2 hold (in fact, both inequalities (i) and (ii) in the second assumptions are satisfied) for $a \leq 1/2$, i.e. if the degree of complementarity in f is sufficiently large.[14]

Lemma 2 *Consider a team with $\theta_1 = \theta_2$. Then if $c'''(x) > 0$ and $\lim_{x \to \overline{x}} c'(x) > \frac{f_1(\bar{x},\bar{x})}{\min\{\theta_1,\theta_2\}}$, effort is interior and identical across agents.*

Suppose that x_1 and x_2 are the optimal effort levels exerted under positive matching by types θ_1 and θ_2 respectively. As outlined in the above example, there are two complications generated by supermodularity in f, relative to our analysis of the additive case. First, whereas the rematching had no output implications in the additive case for a fixed level of input by every team member, now rematching negatively (while keeping efforts unchanged) comes at an output cost of

$$f(x_1, x_1) + f(x_2, x_2) - 2f(x_1, x_2)$$

reflecting the technological benefit of matching like types under technological supermodularity. Second, the first order condition for each team member's choice of effort depends on his partner:

$$x_i = \arg\max_x f(x, x_{i^-})w_i - \theta_i c(x),$$

[14] In general, whichever of the conditions, (i) or (ii), is less restrictive depends on the properties of f. In particular, it is easy to verify that condition (i) more likely to be satisfied than condition (ii) if $\frac{\partial}{\partial x_1}\left(\frac{-1}{f_1(x_1,x_2)}\right) < \frac{\partial}{\partial x_1}\left(\frac{1}{f_2(x_1,x_2)}\right)$ holds, i.e. the effect of the worker's effort impacts his own reward less than the reward of his partner. Conversely, if the reverse holds, (ii) is less restrictive than (i).

implying that

$$w_i = \frac{\theta_i c'(x_i)}{f_1(x_i, x_{i^-})}$$

In the additive case, one could fix the wages in the hypothetical rematching and focus all attention on the frequency of those payments. Now, to keep effort constant, wages will have to change. Here we see the effect that was described above: a lower input partner agrees to exert the same amount of effort for a lower wage when rematched with a higher input partner.

As a result, a sufficient condition for rematching negatively is:

$$\begin{aligned} f(x_1, x_1) + f(x_2, x_2) - 2f(x_1, x_1) \leq & 2\theta_1 c'(x_1) \left[\frac{f(x_1, x_1)}{f_1(x_1, x_1)} - \frac{f(x_1, x_2)}{f_1(x_1, x_2)} \right] + \\ & 2\theta_2 c'(x_2) \left[\frac{f(x_2, x_2)}{f_1(x_2, x_2)} - \frac{f(x_2, x_1)}{f_1(x_2, x_1)} \right] \end{aligned} \tag{6}$$

The left hand side of (6) represents the planner's loss in the expected output under negative matching that occurs due to complementarity if f. The right hand side is the gain from changes in the expected compensation after the switch from positive to negative matching happens.

From the first order conditions for the planner's problem with positive matching (and by invoking the symmetry of effort assignment established in Lemma 2), we have that

$$\theta_i c'(x_i) = \frac{f_1(x_i, x_i)}{2 - f(x_i, x_i) \frac{f_{11}(x_i, x_i) + f_{12}(x_i, x_i)}{(f_1(x_i, x_i))^2} + \frac{f(x_i, x_i)}{f_1(x_i, x_i)} \frac{c''(x_i)}{c'(x_i)}} \tag{7}$$

so we can rewrite the sufficient condition for negative matching as

$$\begin{aligned} f(x_1, x_1) + & f(x_2, x_2) - 2f(x_1, x_2) \leq \\ & \frac{2f_1(x_1, x_1)}{2 - f(x_1, x_1) \frac{f_{11}(x_1, x_1) + f_{12}(x_1, x_1)}{(f_1(x_1, x_1))^2} + \frac{f(x_1, x_1)}{f_1(x_1, x_1)} \frac{c''(x_1)}{c'(x_1)}} \underbrace{\left[\frac{f(x_1, x_1)}{f_1(x_1, x_1)} - \frac{f(x_1, x_2)}{f_1(x_1, x_2)} \right]}_{\Delta(x_1, x_2)} + \\ & \frac{2f_1(x_2, x_2)}{2 - f(x_2, x_2) \frac{f_{11}(x_2, x_2) + f_{12}(x_2, x_2)}{(f_1(x_2, x_2))^2} + \frac{f(x_2, x_2)}{f_1(x_2, x_2)} \frac{c''(x_2)}{c'(x_2)}} \underbrace{\left[\frac{f(x_2, x_2)}{f_1(x_2, x_2)} - \frac{f(x_2, x_1)}{f_1(x_2, x_1)} \right]}_{\Delta(x_2, x_1)} \end{aligned} \tag{8}$$

Notice that this condition nests Proposition 3: when f is additive, (8) boils down to

$$0 \leq \frac{1}{1 + x_1 \frac{c''(x_1)}{c'(x_1)}}(x_1 - x_2) + \frac{1}{1 + x_2 \frac{c''(x_2)}{c'(x_2)}}(x_2 - x_1),$$

and therefore we immediately see that negative matching occurs if $xc''(x)/c'(x)$ is decreasing. Another immediate implication is that when f is log-modular (i.e. Cobb-Douglas), rematching to negative matching for fixed efforts is never optimal because $\Delta(x_1, x_2) = \Delta(x_2, x_1) = 0$. In fact, it is straightforward to verify an even stronger result: if under log-modular f the workers are matched negatively and exert interior efforts, the principal's profit will be higher if he rematches the workers positively while keeping the efforts fixed. Thus, under log-modular f positive assortative matching is optimal.

There is an interesting dichotomy between the cases where f is log-supermodular, and where it is log-submodular. Suppose, without loss, that $x_1 > x_2$. In the case of log-submodularity, $\Delta(x_1, x_2) > 0 > \Delta(x_2, x_1)$, i. e. the expected compensation of the types exerting x_1 falls and that of the types exerting x_2 increases after the matching is switched from positive to negative. As a result, the switch to negative matching is more likely if $c(x)$ is such that $c''(x)/c'(x)$ is decreasing in x sufficiently fast, so that $c''(x_1)/c'(x_1)$ is small and $c''(x_2)/c'(x_2)$ is large, in which case the principal's gain due to a decline in compensation to the workers exerting x_1 will be scaled up, while the loss due to an increase in compensation to the ones exerting x_2 will be scaled down. Note that this observation is consistent with our results in the last section, where we considered a special case of log-submodularity (additive f): for negative matching to be optimal, it was sufficient that $c''(x)/c'(x)$ is decreasing with x faster than $1/x$.

This logic does not, however, carry over to the case where f is log-supermodular. In that case, $\Delta(x_1, x_2) < 0 < \Delta(x_2, x_1)$, i. e. after rematching the workers negatively the principal has to spend more on compensation to those exerting x_l, but saves on the ones exerting x_h. As a result, satisfying inequality (8) is more likely when $c''(x_1)/c'(x_1)$ is very large and $c''(x_2)/c'(x_2)$ is small, i.e. when $c''(x)/c'(x)$ changes with x in exactly the opposite direction in comparison with the log-submodular case. In other words, while an increasing $c''(x)/c'(x)$ led (always) to *positive matching* under additive f (an example of *log-submodularity*), that same property may lead to *negative* matching under *log-supermodularity*. In fact, we can show that there are always examples, for any differentiable, log-supermodular f, such that negative matching is optimal, as long as $c''(x)/c'(x)$ is increasing in x sufficiently fast.

Proposition 4 *Suppose that f is strictly log-supermodular. Then*

(i) For any $\theta_2 > 0$ there exists $\theta_1 < \theta_2$ and increasing convex twice continuously differentiable $c(x)$ such that negative assortative matching is optimal.

(ii) The optimal efforts $x_1 > x_2$ and $\frac{c''(x_1)}{c'(x_1)} > \frac{c''(x_2)}{c'(x_2)}$, where x_1 and x_2 are the optimal effort levels assigned to types θ_1 and θ_2 respectively.

(iii) If $f_1(x, x)$ is weakly decreasing in x, then, under c, if agents were matched positively, those of type θ_1 receive lower reward in case of success.

The first part of the result may seem surprising, because it implies that even a huge amount of technological complementarity may not be enough to drive positive matching. The proof of Proposition 4 is in the Appendix. It relies on the fact that the costs of matching negatively (the left hand side of (8)) are second order when the efforts exerted by different types are nearly equal, given the differentiability assumption. The incentive effects of matching on total compensation (the right hand side of (8)), however, are still first order. Therefore, by choosing $c(x)$ such that $c''(x)/c'(x)$ increases at x_2 sufficiently fast and then picking θ_1 such that, given the chosen $c(x)$, the associated x_1 is sufficiently close to x_2, we isolate cases where the first order incentive effect dominates the second order effect associated with the decline in expected output.

The final part of the result in Proposition 4 shows conditions under which this result exactly reverses the first result about inputs and rewards for modular technologies. Recall that, under modularity, low rewards for high inputs was associated with positive matching. Proposition 4 shows that, with strong complementarities, in contrast to the modular case, high input types receiving low rewards is associated with *negative* matching. Note that the assumption that $f_1(x, x)$ is weakly decreasing is satisfied for constant elasticity functions f.

Another way to think about matching and complementarities in this model is to think about the optimal matching for various degrees of complementarity. For increasing $c''(x)/c'(x)$ we know that positive matching is optimal for small levels of technological complementarity by continuity and the earlier results for the additive case (Proposition 2). Of course, once f becomes very complementary, the cost of negative matching, described by the left hand side, becomes prohibitive. The pattern we expect to emerge for such $c(x)$ is one where, for no or little supermodularity in f matching is positive; then for an intermediate level of supermodularity (in the log-supermodular range) matching is negative, and for extreme levels of supermodularity matching returns to being positive. In the next section we provide an example with constant elasticity of substitution (CES) production function f and a particular cost function c for which such non-monotone relationship between the optimal matching structure and the degree of complementarity in f arises.

4.2 An Example: a CES f and a logarithmic c

In the previous section we showed that for any log-supermodular f it is possible to find such cost function $c(x)$ that negative matching is optimal for some types θ_1 and θ_2. Here we illustrate that such cost function $c(x)$ does not have to be terribly unusual: we show that for the logarithmic cost function and the CES production function negative matching can arise as long as the degree of complementarity in f is high enough.

Suppose f is CES, i.e. $f(x_1, x_2) = (0.5x_1^a + 0.5x_2^a)^{\frac{1}{a}}$, with $a \leq 1$ and $x_i \in [0, 1)$ (to guarantee that $f(x_1, x_2) \in [0, 1)$ for all possible combinations of effort levels). Assumptions 1 and 2 are satisfied for $a \leq 1/2$, thus in the numerical exercises below we will have to check for corner solutions when $a \in (1/2, 1]$. Obviously, the results of Proposition 4 apply for the CES case: if f is log-supermodular (i.e. if $a < 0$), then negative matching would be optimal for some combination of types (θ_1, θ_2) if $\frac{c''(x)}{c'(x)}$ is increasing in x sufficiently fast. However, it turns out that for the CES case, by applying an argument similar to the one used in the proof of Proposition 4, it is possible to obtain a stronger result: instead of claiming that negative matching arises for *some* $c(x)$, we can actually derive a specific condition on a *given* $c(x)$ which guarantees the optimality of negative matching for some combination of types (θ_1, θ_2). This is formally done in the following Proposition:

Proposition 5 *Let* $f(x_1, x_2) = (0.5x_1^a + 0.5x_2^a)^{\frac{1}{a}}$ *and* $a < 0$. *Denote* $h(x) =$

$\frac{1}{1+x\frac{c''(x)}{c'(x)}}$. *Suppose that there exists* $x_2 \in (0,1)$ *such that*

$$\frac{2a}{1-a}h'(x_2)x_2 + 2ah(x_2) > 1. \tag{9}$$

Then for $\theta_2 = \frac{h(x_2)}{4c'(x_2)}$ *and some* θ_1 *sufficiently close to* θ_2 *negative assortative matching is optimal.*

The advantage of condition (9) is that, for a given cost function $c(x)$, one can easily verify whether this condition is satisfied, which allows us to immediately construct a number of examples where the optimal matching structure reverses from positive to negative as the degree of complementarity between the partners' efforts increases.

Consider, for example, the cost function $c(x) = -\ln(1-x)$. Here $\frac{c''(x)}{c'(x)} = \frac{1}{1-x}$ is increasing in x, and hence in the modular case the optimal matching is positive, independently of the agents' types (provided that the effort assignments are interior). However, for supermodular f optimal matching structure can change to negative: since $h(x) = 1-x$, condition (9) is equivalent (for $a < 0$) to

$$x_2 > \frac{2a-1}{2a} \cdot \frac{1-a}{2-a} \tag{10}$$

This condition is satisfied for some $x_2 \in (0,1)$ as long as $\frac{2a-1}{2a} \cdot \frac{1-a}{2-a} < 1$, which, for $a < 0$, is equivalent to $a < -1$. As a result, for any $a < -1$, there exists a set of types (θ_2, θ_1) such that negative matching is optimal. More specifically, by employing the first order condition for the planner's problem under positive matching, negative matching is optimal for all $\theta_2 < \left(1 - \frac{2a-1}{2a} \cdot \frac{1-a}{2-a}\right)^2/4$ (so that (10) holds at the optimal x_2) and all θ_1 sufficiently close to such θ_2.

Next, observe that the range of θ_2 for which (10) holds at the optimal x_2 is not monotone: as a decreases from -1 below, the upper bound on such θ_2 first rises and then falls to 0 as a becomes arbitrary small. This might serve as an indication that, for a given combination of types (θ_1, θ_2), as a falls from 1 to $-\infty$, the optimal matching structure switches from positive (since positive matching is optimal for $a = 1$ by Proposition 2) to negative (if θ_1 and θ_2 are chosen in accordance with Proposition 5) and then back to positive (when the degree of complementarity becomes very large), as we hypothesized in the end of section

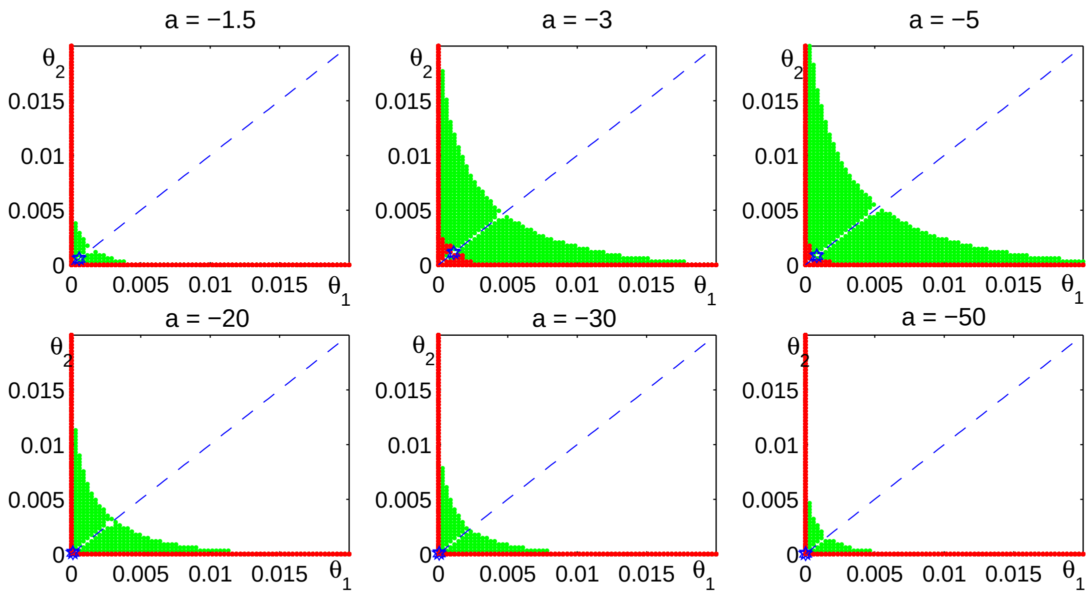

Figure 1: Ranges of (θ_1, θ_2) for which negative matching is optimal, $f(x_1, x_2) = (0.5x_1^a + 0.5x_2^a)^{1/a}$ and $c(x) = -\ln(1-x)$.

4.1. However, the condition in Proposition 5 is sufficient but not necessary: in particular, it limits the benefits from switching to negative matching by analyzing only the case when the effort levels do not adjust in response to the matching structure. Therefore, further analysis is needed to confirm that for large degrees of complementarity optimal matching indeed reverses back to positive, as well as to understand whether the possibility of effort adjustment after switching from positive to negative matching has a significant impact on the matching structure. There exists no analytical solution for the optimal effort levels under negative matching structure, thus we continue with numerical analysis.

Shaded areas on Figure 1 illustrate the ranges of (θ_1, θ_2) for different values of a for which negative matching is optimal. Each of the plots decomposes the shaded area into two: a red (darker) area is the set of parameters for which the switch from positive to negative matching is optimal even without effort adjustment, and a green (lighter) area is the one where efforts must adjust to make the switch from the positive to the negative matching beneficial, while the single blue star on the diagonal corresponds to the maximum θ_2 for which condition (10) is satisfied at the optimal x_2 (and hence switching from positive to negative matching without effort adjustment is optimal for combinations of (θ_1, θ_2) in the neighborhood of the blue star).

First, our numerical results show that no negative matching arises for $a \in [-1, 1]$ (thus we do not show any plots for a from this range), but, consistently with the analytical results above, when $a < -1$ negative matching is optimal for some combinations of (θ_1, θ_2). Notice that even for very large degrees of complementarity (e.g. $a = -30$) the range of parameters leading to negative matching is non-trivial. Second, it turns out that the possibility of adjusting effort levels after switching from positive to negative matching considerably broadens the set of parameters for which negative matching arises: even though Propositions 4 and 5 serve our main purpose by illustrating that moral hazard can lead to negative matching even in the presence of large technological complementarities, they only describe a narrow class of situations in which this might be happening. Third, Figure 1 confirms that the optimal matching structure is not monotone in a: as the degree of complementarity increases, the range of (θ_1, θ_2) for which negative matching is chosen first broadens and then shrinks, implying that there exist combinations of types for which lowering a induces switching from positive to negative and then back to positive matching structure.

4.3 Endogenous Team Formation

With supermodular production function f we can ask an additional question: when do teams form despite the difficulties generated by the moral hazard problem? This question is interesting because, generally, for a model to generate team production endogenously, it must build in complementarities in the production technology. Complementarities generally drive positive matching. Here, however, since we have regions where the link between technological complementarity and positive matching is broken, there is a possibility that negatively matched teams appear endogenously.

To put it more formally, our goal is to understand when the profit from forming the teams (and matching the team members optimally) is greater than the profit from making each worker operate independently (and thus avoiding the moral hazard problem). Under the CES production technology, a partial answer to the question of the endogenous formation of team is straightforward. If a worker

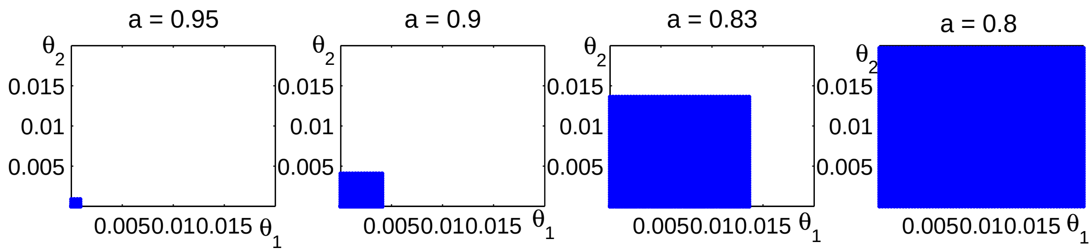

Figure 2: Ranges of (θ_1, θ_2) for which forming teams is optimal, $f(x_1, x_2) = (0.5x_1^a + 0.5x_2^a)^{1/a}$ and $c(x) = -\ln(1-x)$.

operates without a partner, the probability of high output is given by $f(x, 0)$, where x is the effort exerted by the worker, i.e. it is equivalent to saying that the worker's partner exerts zero effort. When f is CES and $a \leq 0$, $f(x, 0) = 0$ for any x, implying that it is never optimal to split a team into two independent workers. Thus, for all $a \leq 0$, the planner optimally chooses to form teams, and for some (θ_1, θ_2) (as established in Proposition 4 and illustrated on Figure 1 for a logarithmic cost function) it is optimal to sort these teams negatively.

When $a \in (0, 1]$, $f(x, 0) > 0$ and the planner may choose not to form teams in order to avoid the moral hazard cost. Obviously, this is more likely to happen for large a, for which the gains in expected output due to complementarity in the underlying technology are small. In particularly, if $a = 1$, no teams are formed. Figure 2 illustrates the ranges of (θ_1, θ_2) for which team formation is optimal (and, since in our example negative matching happens to be optimal only for $a < -1$, all the teams formed under the parameters on Figure 2 are positive). Naturally, as a falls, these sets expand: for a close to 1 teams are formed only if the costs are very small, because this is where optimal effort levels are relatively high and the gains from exploring the complementarity in the underlying technology are large enough to overcome the moral hazard costs; as a falls, higher degree of complementarity induces team formation even if the agents' optimal effort levels are not extremely large.

This exercise further underscores the basic message of this section: moral hazard can turn around the conventional connection between the shape of the technology and the outcomes. Here, we see that a technology can have sufficient complementarities for it to be optimal to have agents work together, and yet the optimal matching structure is negative. Such an outcome is impossible in the conventional model with no information frictions, since technological complemetarity immediate generates positive matching.

5 Conclusion

We show that information frictions, in the form of a moral hazard in teams problem, can have an impact on matching predictions. In the case where there is no matching prediction under full information, we show that there is typically a non-trivial matching decision for the principal. We describe the channel by which the shape of the technology and the matching structure are linked. When we move to a world with complementarities in production, a further effect is felt through the required reward to get a given level of effort. We use this to show that the matching structure may be non-monotone in technological complementarities: sometimes greater underlying complementarities move the matching structure from positive to negative, the opposite of the usual logic under complete information. The link between observed matching structure and the complementarity of the underlying technology is completely broken. Finally, we show that this force can allow the model to explain team production endogenously through complementarities, while still predicting negatively matched teams.

References

[1] A.A. Alchian, H. Demsetz, Production, information costs, and economic organization, Amer. Econ. Rev., 62 (1972) 777-795.

[2] A. Anderson, L. Smith, Dynamic matching and evolving reputations, Rev. Econ. Stud., 77 (2010) 3-29.

[3] G. Becker, A theory of marriage: Part I, J. Polit. Econ., 81 (1973) 813-846.

[4] A. Franco, M. Mitchell and G. Vereshchagina, Incentives and the structure of teams, 2009, mimeo.

[5] 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, J. Polit. Econ., 111 (2003) 465-497.

[6] B. Holmstrom, Moral hazard in teams, Bell J. Econ., 13 (1982) 324-340.

[7] S. Jeon, Moral hazard and reputational concerns in teams: Implications for organizational choice, Int. J. Ind. Org., 14 (1996) 297-315.

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

[9] A. Kaya, G. Vereshchagina, Moral hazard and sorting in a market for partnerships, Working Paper, Arizona State University, 2009.

[10] M. Kremer, The O-ring theory of economic development, Quart. J. Econ., 108 (1993) 551-575.

[11] P. Legros, A.F. Newman, Monotone matching in perfect and imperfect worlds, Rev. Econ. Stud., 69 (2002) 925-942.

[12] P. Legros, A.F. Newman, Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities, Econometrica, 75 (2007) 1073 - 1102.

[13] R.P. McAfee, J. McMillan, Optimal contracts for teams, Intern. Econ. Rev., 32 (1991) 561-577.

[14] M.A. Meyer, The dynamics of learning with team production: Implications for organizational choice, Quart. J. Econ., 109 (1994) 1157-1184.

[15] A.F. Newman, Risk-bearing and entrepreneurship, J. Econ. Theory, 137 (2007) 11-26.

[16] J.A. Nickerson, B.S. Silverman, Why aren't truck drivers owner-operators? Asset ownership and the employment relation in interstate for-hire trucking, J. Econ. Manage. Strategy, 12 (2003) 91-118.

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

[18] H. Thiele, A. Wambach, Wealth effects in the principal agent model, J. Econ. Theory, 89 (1999) 247-260.

6 Appendix

Proof of Proposition 1. Suppose to contradiction 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.

The total surplus of the firm owner matching his teams positively is equal to

$$\Pi^{+} = 2x_1(1 - 2w_1) + 2x_2(1 - 2w_2).$$

As was noted above, the same contracts (x_1, w_1) and (x_2, w_2) remain 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 contracts, the firm owner would obtain profit

$$\Pi^{-} = 2(x_1 + x_2)(1 - w_1 - w_2).$$

It is easy to verify that

$$\Pi^{-} - \Pi^{+} = 2(x_1 - x_2)(w_1 - w_2),$$

which is positive if high-input workers receive high rewards. Hence, if the high-input types receive high compensation, 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. This completes the proof of (i).

The proof of (ii) is based on a symmetric argument. ■

Proof of Lemma 1. The planner's profit from a team consisting of two workers with θ_1 and θ_2 (possibly $\theta_1 = \theta_2$)is given by

$$\Pi(x_1, x_2) = (x_1 + x_2)\left(1 - \theta_1 c'(x_1) - \theta_2 c'(x_2)\right). \qquad (11)$$

Note, first, that $c'(0) < 1/\min\{\theta_1, \theta_2\}$ implies that both workers cannot optimally have $x = 0$ (since under this condition profits are clearly positive for a small reward offered to one worker, while the profit would be zero if both workers receive zero compensation, and exert no effort). Further, $c'(\overline{x}) > 1/\min\{\theta_1, \theta_2\}$ implies that none of the workers will be asked to exert $x = \overline{x}$ (otherwise the planner's profit will be negative). Therefore, the only possible corner solution that remains to be eliminated is one where (without loss) worker 1 works an interior amount and worker 2 exerts $x_2 = 0$. In this case, the first order condition with respect to x_1 holds with equality:

$$1 - \theta_1 c'(x_1) - \theta_2 c'(x_2) = (x_1 + x_2)\theta_1 c''(x_1) \tag{12}$$

The firm's marginal return to effort by worker 2 is

$$1 - \theta_1 c'(x_1) - \theta_2 c'(x_2) - (x_1 + x_2)\theta_2 c''(x_2)$$

which is equal to

$$(x_1 + x_2)(\theta_1 c''(x_1) - \theta_2 c''(x_2))$$

If $x_2 = 0$, this is strictly positive if $c''(0) = 0$ and $c''(x) > 0$ for all x, implying that (11) is maximized at some $x_2 > 0$. Therefore, it would be optimal to have both workers exert strictly positive effort if $c'(0) = 0$ were satisfied. When $c'(0) > 0$, there is a fixed cost of making a worker exert even an arbitrary small amount of effort, which could potentially create incentives for the planner to shut one of the workers down and pay him no wage at all.[15] However, it is straightforward to verify that, by continuity, for $c'(0)$ sufficiently small (i.e. if $c'(o) < \varepsilon$ for some $\varepsilon > 0$), doing this would be suboptimal and the interior solution will be chosen even in the presence of fixed costs. this completes the proof of interiority in (ii).

Observe also that if $\theta_1 = \theta_2$ then the marginal product of worker 2 at $x_2 = 0$ is strictly positive as long as $c''(x_1) > c''(0)$, i.e. assuming that $c'''(x) > 0$ for all $x \in [0, \overline{x}]$ eliminates the possibility of the corner solution in question for the positively matched teams, and assumption $c''(0) = 0$ becomes unnecessary.

Additionally, for a team with $\theta_1 = \theta_2$, it is immediate that equating marginal products in the interior solution implies

$$c''(x_1) = c''(x_2)$$

[15] If the principal does so, (11) modifies to $x_1(1 - \theta_1 c'(x_1))$, i.e. the fixed cost associated with making agent 2 work a positive amount is avoided.

so that $c''' > 0$ implies $x_1 = x_2$. Therefore, the solutions for positive teams are symmetric.

Finally, to prove uniqueness, we need to check the second order conditions. We have

$$\frac{\partial^2 \Pi(x_1, x_2)}{\partial x_1^2} = -2\theta_1 c''(x_1) - \theta_1 (x_1 + x_2) c'''(x_1) \tag{13}$$

and symmetrically for x_2. These are negative since c'' and c''' are both positive. The cross-derivative is

$$\frac{\partial^2 \Pi(x_1, x_2)}{\partial x_1 \partial x_2} = -\theta_1 c''(x_1) - \theta_2 c''(x_2), \tag{14}$$

Note that by the first order conditions,

$$\theta_1 c''(x_1) = \theta_2 c''(x_2) = (1 - \theta_1 c'(x_1) - \theta_2 c'(x_2)) / (x_1 + x_2)$$

This implies that the determinant is positive since the on-diagonal elements are bigger in absolute value. Note that for the latter part, $c''' > 0$ is necessary. ■

Proof of Proposition 3. For a team of two identical agents of type θ, the first order conditions (12) can be simplified to

$$\frac{1}{\theta c'(x)} = 2\left(1 + \frac{x c''(x)}{c'(x)}\right), \tag{15}$$

where the left hand side is the inverse of the compensation paid after the realization of high output. If $\frac{x c''(x)}{c'(x)}$ is decreasing, it follows from (15) that higher inputs receive higher compensation, which, by Proposition 1, contradicts the optimality of positive matching. ■

Proof of Lemma 2. The planner's profit from a team (θ_1, θ_2) is

$$\max_{x_1, x_2} f(x_1, x_2) \cdot \left(1 - \frac{\theta_1 c'(x_1)}{f_1(x_1, x_2)} - \frac{\theta_2 c'(x_2)}{f_2(x_1, x_2)}\right). \tag{16}$$

Obviously, $\lim_{x\to\bar{x}} c'(x) > \frac{f_1(\bar{x},\bar{x})}{\min\{\theta_1,\theta_2\}}$ implies that $x_i < \bar{x}$ (otherwise the firm's profit would be negative). The firm's marginal return to effort by worker 1 is

$$f_1(x_1,x_2)\cdot\left(1-\frac{\theta_1 c'(x_1)}{f_1(x_1,x_2)}-\frac{\theta_2 c'(x_2)}{f_2(x_1,x_2)}\right)$$
$$+\theta_1 f(x_1,x_2)\frac{f_{11}(x_1,x_2)c'(x_1)-c''(x_1)f_1(x_1,x_2)}{(f_1(x_1,x_2))^2}+\theta_2 f(x_1,x_2)\frac{f_{21}(x_1,x_2)c'(x_2)}{(f_2(x_1,x_2))^2},$$

and the marginal return to effort by worker 2 is symmetric.
Assuming $\lim_{x_1\to 0} f_1(x_1,x_2)=\infty$ and $\lim_{x_2\to 0} f_2(x_1,x_2)=\infty$ guarantees that the marginal products are strictly positive at zero, implying that the solution must be interior.

Next, we verify that, if workers are sorted positively, the efforts within a team are identical. Let θ be the types of the workers in a team. Suppose that the effort assignment is not symmetric, with $x_1 > x_2 \geq 0$.

Consider perturbing x_1 and x_2 by ε, decreasing the first and increasing the second. We show that, under condition (i) of Assumption 2 such perturbation always increases the payoff for the planner. The payoff after the perturbation is

$$f(x_1-\varepsilon,x_2+\varepsilon)\left(1-\frac{\theta c'(x_1-\varepsilon)}{f_1(x_1-\varepsilon,x_2+\varepsilon)}-\frac{\theta c'(x_2+\varepsilon)}{f_2(x_1-\varepsilon,x_2+\varepsilon)}\right)$$

Since f is supermodular, such perturbation (weakly) increases f, so we will focus on the impact on the payoff term only. Taking the derivative with respect to ε and evaluating at $\varepsilon=0$, the decline in the payoff can be divided into three parts. The first is

$$\frac{\theta c''(x_1)}{f_1(x_1,x_2)}-\frac{\theta c''(x_2)}{f_2(x_1,x_2)}$$

Concavity and supermodularity of f, together with $c''' > 0$, imply that this term is strictly positive. The second term is

$$\theta c'(x_1)\frac{f_{12}(x_1,x_2)-f_{11}(x_1,x_2)}{f_1(x_1,x_2)^2}-\theta c'(x_2)\frac{f_{12}(x_1,x_2)-f_{22}(x_1,x_2)}{f_2(x_1,x_2)^2},$$

which, by convexity of $c(x)$ and condition (i) of Assumption 2 is also positive. Therefore, the considered perturbation unambiguously increases the planner's expected payoff.

To verify that condition (ii) of Assumption 2 independently implies that the effort choices in a team of homogeneous agents are symmetric, consider an alternative perturbation: increase x_2 by ε and reduce x_1 by $\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)}\varepsilon$. For ε converging to 0, such perturbation leaves $f(x_1, x_2)$ unchanged while reducing the payoff term by

$$\theta c'(x_1)\frac{f_{12}(x_1,x_2) - f_{11}(x_1,x_2)\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)}}{f_1(x_1,x_2)^2} - \theta c'(x_2)\frac{f_{12}(x_1,x_2)\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)} - f_{22}(x_1,x_2)}{f_2(x_1,x_2)^2},$$

which can be rearranged as

$$\frac{\theta c'(x_1)}{f_1(x_1,x_2)}\frac{f_{12}(x_1,x_2) - f_{11}(x_1,x_2)\frac{f_2(x_1,x_2)}{f_1(x_1,x_2)}}{f_1(x_1,x_2)} - \frac{\theta c'(x_2)}{f_1(x_1,x_2)}\frac{f_{12}(x_1,x_2) - f_{22}(x_1,x_2)\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}}{f_2(x_1,x_2)}$$

Condition (ii) of Assumption 2, together with convexity of $c(x)$, imply that the this term is positive, and hence the payoff of the planner increases unambiguously. ■

Proof of Proposition 4. Our proof is constructive. We begin with some $\tilde{c}$ with associated x_h chosen for a positively matched team of type θ_h, then show that for some x_l and modified cost function c, (8) is satisfied. Finally, we show that x_l is optimal for some θ_l under our modified cost function.

In particular, suppose that, for $\tilde{c}(x)$,

$$\frac{f_1(x,x)}{2 - f(x,x)\frac{f_{11}(x,x)+f_{12}(x,x)}{(f_1(x,x))^2} + \frac{f(x,x)}{f_1(x,x)}\frac{\tilde{c}''(x)}{\tilde{c}'(x)}}$$

is decreasing in x. This guarantees that the FOC

$$\theta\tilde{c}'(x) = \frac{f_1(x,x)}{2 - f(x,x)\frac{f_{11}(x,x)+f_{12}(x,x)}{(f_1(x,x))^2} + \frac{f(x,x)}{f_1(x,x)}\frac{\tilde{c}''(x)}{\tilde{c}'(x)}} \tag{17}$$

has a unique solution, and hence it uniquely determines the planner's optimal effort allocation.Our modification alters $\tilde{c}$ only at points beyond x_h. In particular, consider

$$g(x) = \begin{cases} x & x < x_h \\ \frac{1}{A}\sinh(A(x - x_h)) + x_h & x \geq x_h, \end{cases}$$

where $\sinh(x) = \left(e^x - e^{-x}\right)/2$. Note that $g(x)$ is twice continuously differentiable, increasing, weakly convex, and has a positive third derivative. Moreover,

$$g''(x)/g'(x) = \begin{cases} 0 & x < x_h \\ A\tanh(A(x - x_h)) & x \geq x_h, \end{cases}$$

so g''/g' is weakly increasing. As a result, $c(x) = \tilde{c}(g(x))$ satisfies all of the assumptions, and, since the right hand side of (17) is decreasing even faster under $c(x)$ than under $\tilde{c}(x)$, there is still a unique solution.

Next we construct x_l and choose the parameter A such that, if x_l and x_h are the optimal efforts under positive matching, then rematching negatively would be optimal since (8) is satisfied. When f is strictly log-supermodular, $\frac{f(x_1,x_2)}{f_1(x_1,x_2)}$ is strictly decreasing in x_2. Therefore, $\Delta(x_l, x_h) < 0$ and $\Delta(x_h, x_l) > 0$ for all $x_l > x_h$, where $\Delta(\cdot,\cdot)$ is defined in (8). In addition, $\Delta(x_h, x_h) = 0$ and $\Delta_2(x_h, x_h) > 0$ (by strict log-supermodularity of f), while both the left hand side of (8) and its derivative with respect to x_l are equal to zero at $x_l = x_h$. Thus there exists $\varepsilon > 0$ and $x_l > x_h$ such that

$$f(x_l, x_l) + f(x_h, x_h) - 2f(x_l, x_h) < \frac{2f_1(x_h, x_h)}{2 - f(x_h, x_h)\frac{f_{11}(x_h,x_h)+f_{12}(x_h,x_h)}{(f_1(x_h,x_h))^2} + \frac{f(x_h,x_h)}{f_1(x_h,x_h)}\frac{c''(x_h)}{c'(x_h)}} \cdot \Delta(x_h, x_l) - \varepsilon \tag{18}$$

Further, since $\frac{f(x,x)}{f_1(x,x)} > 0$ for all $x \in (0, \overline{x})$, there exists $\overline{r} > 0$ such that for all $r > \overline{r}$

$$\frac{2f_1(x_l, x_l)}{2 - f(x_l, x_l)\frac{f_{11}(x_l,x_l)+f_{12}(x_l,x_l)}{(f_1(x_l,x_l))^2} + \frac{f(x_l,x_l)}{f_1(x_l,x_l)} \cdot r} \cdot \Delta(x_l, x_h) > -\varepsilon \tag{19}$$

Obviously, (18) and (19) imply that if $\frac{c''(x_l)}{c'(x_l)} \geq \overline{r}$ holds, inequality (8) would be satisfied. Observe that, for any x_l, by increasing A in $g(x)$, we can make $\frac{c''(x_l)}{c'(x_l)} \geq \overline{r}$. This determines the choice of A.

Hence, all that remains is to show that the relevant x_l is optimal for some θ_l, so that our example is constructed. However, the set of maximizers $x_l(\theta_l)$ is a continuous function (by the theorem of the maximum), and grows large as θ_l goes to zero. Therefore there must exist $\theta_l < \theta_h$ such that x_l is optimal for θ_l. ■

Proof of Proposition 5. Under the CES f, sufficient condition (8) for negative matching reduces to

$$x_1 + x_2 - 2\left(\frac{1}{2}x_1^a + \frac{1}{2}x_2^a\right)^{\frac{1}{a}} \leq 2\theta_1 c'(x_1)\left(x_1 - x_1^{1-a}x_2^a\right) + 2\theta_2 c'(x_2)\left(x_2 - x_2^{1-a}x_1^a\right), \tag{20}$$

where, from the first order conditions for the planner's problem,

$$2\theta_i c'(x_i) = \frac{1/2}{1 + x_i \frac{c''(x_i)}{c'(x_i)}}. \tag{21}$$

Substitute (21) into (20) and denote by $D(x_1, x_2)$ the difference between the right hand side and the left hand side of (20). We need to find conditions under which $D(x_1, x_2) > 0$, i.e negative matching is strictly optimal. It is straightforward to verify that $D(x_2, x_2) = 0$ and $D_1(x_2, x_2) = 0$ for any x_2. Thus, $D(x_1, x_2)$ would be positive for x_1 sufficiently close to x_2 if $D_{11}(x_2, x_2) > 0$. By differentiating $D(x_1, x_2)$ twice, we obtain

$$D_{11}(x_2, x_2) = 2ah'(x_2) + 2a(1-a)\frac{h(x_2)}{x_2} - (1-a)\frac{1}{x_2},$$

which is positive if (9) holds. Thus, if there exists x_2 such that (9) is satisfied, and x_2 and some x_1 sufficiently close to it are the optimal effort levels under positive matching, then switching to negative matching while preserving the same effort allocation is optimal. Since f satisfies Assumptions 1 and 2 for $a \leq 1/2$, by the first order conditions $\theta_2 = \frac{h(x_2)}{4c'(x_2)}$ is the type for which x_2 is optimal. By continuity, θ_1 must be chosen sufficiently close to θ_2. ■