BANCO CENTRAL DE COSTA RICA

DOCUMENTO DE TRABAJO
N.º 03|2021

# The Optimal Assortativity of Teams Inside the Firm

Ashwin Kambhampati
Carlos Segura-Rodriguez

Cover photo: “Presentes”, a bronze sculpture set, year 1983, by the Costa Rican artist Fernando Calvo Sánchez. Collection of the Central Bank of Costa Rica.

# The Optimal Assortativity of Teams Inside the Firm*

Ashwin Kambhampati† Carlos Segura-Rodriguez‡

September 17, 2020



## Abstract

How does a profit-maximizing manager form teams and compensate workers in the presence of both adverse selection and moral hazard? Under complete information, it is well known that any complementarity in characteristics implies that positive assortative matching is productively efficient. But, under asymmetric information, we uncover the problem of *disassortative incentives*: incentive costs may increase in assortativity. Profit maximization thus prescribes either random or negative assortative matching, both productively inefficient, when complementarities are weak and effort costs are high enough. When this is the case, the manager may instead prefer to delegate matching, allowing workers to sort themselves into teams. Our results shed light on recent empirical work documenting patterns of non-assortative matching inside of firms.

**Key words:** Asymmetric Information, Assortative Matching, Delegation, Teams.

**JEL Codes:** C78, D86,L23.

*We thank George Mailath, our advisor, and Aislinn Bohren, Ralph Boleslavsky, Rohit Lamba, Annie Liang, Andy Postlewaite, Steven Matthews, Lones Smith, Juuso Toikka, and Rakesh Vohra for helpful comments and suggestions.

†University of Pennsylvania, Department of Economics, akambh@sas.upenn.edu

‡Central Bank of Costa Rica, Research Department, segurarc@bccr.fi.cr.

# Contents

# 1 Introduction

Teamwork has increasingly become “a way of life" in many firms (Lazear y Shaw, 2007). For instance, at Google, “the team is the molecular unit where real production happens, where innovative ideas are conceived and tested, and where employees experience most of their work."[1] Yet, forming teams composed of complementary and productive workers is complicated. First, workers often possess private information about their characteristics, such as their ability or willingness to work collaboratively. Second, individual effort is difficult to identify from team output. A profit-maximizing manager must therefore design contracts that simultaneously screen for unobservable characteristics and provide incentives for effort in teams.

What is the optimal assignment of workers to teams? How should a manager remunerate her workers? With few exceptions, the economic theory of teams has answered each question separately. To wit, a large literature in matching, pioneered by Becker (1973), studies the optimal composition of teams, abstracting from incentives. At the same time, a large literature in contract theory, pioneered by Hölmstrom (1982), studies the provision of incentives within a single team, fixing its composition.[2]

By conducting a unified analysis of optimal team composition and incentives, we uncover a novel economic distortion: Even when matching likes with likes– positive assortative matching (PAM)– is productively efficient, creating the right incentives for workers to truthfully reveal their characteristics and exert effort can make implementing PAM prohibitively costly, leading a profit-maximizing manager to match non-assortatively. We identify when and why productive distortions occur, and argue that non-assortative matching within a firm need not indicate a lack of productive complementarity. Moreover, we

[1] https://rework.withgoogle.com/guides/understanding-team-effectiveness/steps/introduction/

[2] Less related is the work of Marschak y Radner (1972), which investigates the behavior of a fixed team of agents whom share a common prior and objective function, but possess different information when taking actions. Within this framework, Prat (2002) studies the optimal composition of teams.

identify when a profit-maximizing manager would prefer to allow workers to sort themselves in order to save on incentive costs. Together, our results rationalize recent empirical evidence of non-assortative matching inside of firms (Adhvaryu et al. (2019)).

We posit a simple model to illustrate the key mechanism. A single risk-neutral manager assigns risk-neutral workers, each protected by limited liability, to teams of two. She then compensates workers upon observing the output that each team produces. Each worker has a private type, high or low, and can exert costly hidden effort. Effort by both teammates is necessary to produce high output with positive probability; high workers ("highs") are more productive than low workers ("lows"); and there are complementarities between types–the productivity gain from working with a high is strictly increasing in own type (the production technology is supermodular).[3]

We characterize optimal wages in terms of the assortativity of the matching the manager implements and the production function. A matching exhibits positive assortativity if highs are more likely than lows to match highs. The production technology is log submodular if the proportionate gain from matching a high is decreasing in own type. Theorem 1 states that highs and lows must both receive strictly positive information rent if, and only if, the implemented matching exhibits strict positive (negative) assortativity and the production technology is strictly log submodular (supermodular).

The result is a consequence of two different incentive problems. The first is standard. In order to incentivize a low to exert effort, she must pay him a high enough wage for producing high output relative to what she pays him for producing low output. But, highs are more likely than lows to produce high output. Providing effort incentives for lows thereby increases the payoff to highs from misreporting their type. To dissuade such deviations, the manager must always pay highs a strictly positive information rent.

[3] Interestingly, all distortions we identify hold *only* when production is supermodular. In particular, if production is strictly submodular, Theorem 2 directly implies that there is no incentive-efficiency tradeoff.

The second is non-standard–we call it the problem of *disassortative incentives*. When lows are held to their reservation value (so that they receive zero information rent), the payoff a high receives after deviating is determined by the proportionate gain he gets from being a high, rather than a low, in his assigned team. But if the production technology is strictly log submodular, then this gain is strictly higher when he is assigned to work with a *low* instead of a high. Counterintuitively, deviation payoffs are therefore *increasing* in the assortativity of the implemented matching. Increasing wages for highs to satisfy the *downward* incentive constraint, however, makes *upward* deviations by lows profitable. Satisfying both incentive constraints thus requires paying lows a strictly positive information rent.

Disassortative incentives give rise to a novel rent-efficiency tradeoff: If the production technology is strictly log submodular, optimal wage payments *strictly increase* in the assortativity of the matching the manager implements. Theorem 2, our main result, shows that PAM is suboptimal if and only if effort costs lie above a positive threshold. Moreover, as talent becomes scarce or abundant (the proportion of highs in the population approaches zero or one), this threshold approaches zero, so that PAM is suboptimal for *all* cost parameters. In the former case, random matching (RM) is optimal while in the latter negative assortative matching (NAM) is optimal.[4]

To conclude our analysis, we consider the implications of our result for the internal organization of the firm. We ask the following question: If a manager could commit to *not* asking workers to report their types, so that she would instead have to delegate the problem of sorting to her workers, would she do so? On one hand, such an arrangement entails a *loss of control*: She can no longer tailor wages to reported characteristics. But on the other, the manager

[4]The intuition for the optimality of RM is subtle, but follows directly from our discussion of Theorem 1: If the production technology is strictly log submodular, then as the implemented matching goes from exhibiting positive assortativity to exhibiting negative assortativity, the manager goes from paying information rents to both highs and lows, to paying rents to highs alone. Hence, the manager saves less on incentive costs when distorting the matching past the point at which the matching is random, i.e. incentive costs have a kink. If the cost of effort is not too high, in which case NAM is optimal, and not too low, in which case PAM is optimal, then it turns out that RM is optimal.

can exploit her workers' local information about each other's characteristics. We formalize this tradeoff by considering an environment in which there is no reporting stage, but in which workers are endowed with knowledge of one another's types. Using this knowledge, workers then form self-enforcing teams. Theorem 3, our final result, shows that if the firm is large enough, talent is scarce, and effort is not too costly, then delegation is optimal. We thus provide a theoretical rationale for the increasing prominence of self-organized teams within firms (see Kambhampati et al. (2018) and the references therein).

New empirical evidence supports our theoretical results. Using three years of daily data on worker-level productivity and team composition in a large Indian garment manufacturer, Adhvaryu et al. (2019) find evidence for NAM despite production complementarities between workers.[5] They hypothesize that NAM arises due to large negative consequences of failing to meet deadlines to complete and deliver an order, i.e. frayed buyer-supplier relations. Our results show that NAM can be rationalized even in the absence of such considerations, as long as complementarities are sufficiently low. Quantifying the relative importance of each channel is therefore a promising avenue for future empirical research.[6]

### 1.1 Literature

We summarize the closest related theoretical literature. Franco et al. (2011) and Kaya y Vereshchagina (2014) consider settings in which a profit-maximizing manager assigns workers to teams subject to moral hazard. Our model enriches these frameworks to include adverse selection. Damiano y Li (2007) and Johnson (2013) find conditions for PAM to be profit-maximizing in en-

[5] Aggregate output would increase by an estimated 1%- 4% if the firm switched to PAM. We note that Theorem 3 implies that whenever NAM is optimal under centralized matching delegation is suboptimal.

[6] Understanding whether productivity losses are driven by supply chain constraints, as hypothesized by Adhvaryu et al. (2019), or by the degree of technological complementarity in the firm, as our theory suggests, is of considerable importance for development policy. In particular, Adhvaryu et al. (2019) suggest that reducing the market power of large multinational buyers might result in lower inequality between firms in developing and developed countries. In contrast, our theory suggests a "tipping point" explanation, unrelated to market power, wherein the endogenously chosen productivity of a firm is discontinuous in the degree of technological complementarity.

vironments in which individuals have private information, but match payoffs are specified exogenously.[7] Our model enriches these frameworks to include moral hazard within teams. To distinguish our contribution from these papers, we assume a production specification ensuring that neither moral hazard nor adverse selection alone generates a distortion of PAM.[8] Indeed, all of our results are driven by the interaction between the two.

More broadly, building on Becker (1973)'s marriage model, a number of papers have investigated the role of imperfectly transferable utility and costly search in distorting the stability of PAM.[9] Though we share a similar motivation as these papers, our analysis is quite different due to our focus on profit-maximization under asymmetric information rather than stability under complete information. Nevertheless, the sufficient condition for PAM we identify, log supermodularity, is identical to the one found by Smith (2006), who studies the problem of random search for marriage partners in an environment with nontransferable utility. In Smith (2006)'s theory, the role of this condition is to ensure that the opportunity cost of time is small enough relative to the reward of finding a higher quality match. In our theory, the condition ensures that the gain a high worker obtains from misreporting his type is decreasing in the assortativity of the matching the manager implements.

Our modeling of the moral hazard in teams problem follows the literature in which limited liability constraints are the source of contracting friction (in addition to Franco et al. (2011) and Kaya y Vereshchagina (2014), see, among

[7] The problem of a profit-maximizing platform also bears resemblance to the one faced by the manager in our study. See, for instance, Gomes y Pavan (2016) and the references therein.

[8] Precisely, our specification, which nests Kremer (1993)'s O-Ring model as a special case, implies that type and effort are complements. As shown by Franco et al. (2011) and Kaya y Vereshchagina (2014), this leads to PAM absent adverse selection.

[9] For instance, Legros y Newman (2007) find conditions for PAM in general environments with imperfectly transferable utility; Serfes (2005) and Serfes (2007) find conditions for PAM when principals match agents (see also Wright (2004)); Shimer y Smith (2000) find conditions for PAM when individuals engage in random search; and Eeckhout y Kircher (2010) find conditions for PAM when individuals engage in directed search. An interesting, recent application of the framework of Legros y Newman (2007) is Chiappori y Reny (2016), who find NAM is optimal when individuals are heterogeneous in risk aversion and match to share risk. See also Anderson y Smith (2010), who find NAM to be optimal in a dynamic sorting environment in which workers have a high discount factor and possess a public reputation, and Chade y Eeckhout (2018), who find a natural informational environment in which the match surplus generated by a team is submodular in characteristics, so that NAM is optimal.

many other papers, Sappington (1983), Sappington (1984), Innes (1990), and Che y Yoo (2001)). This ensures that, despite our assumption of risk neutrality, “efficiency wages" must be paid to incentivize effort. Outside of this literature, McAfee y McMillan (1991) consider the interaction of adverse selection and moral hazard within a fixed team in the absence of limited liability constraints. They establish conditions under which incentives are linear in team output, even when individual performance measures are available.

Finally, our modeling of the tradeoff between centralized assignment and delegated matching is inspired by the literature on delegation, i.e. Aghion y Tirole (1997) and Dessein (2002), wherein the benefit of delegation is that workers can utilize superior information to make decisions and the cost is a loss of control. In contrast to these theories, however, the loss of control issue in our model is not related to the misalignment of incentives between workers and the manager. Indeed, under strictly increasing differences, endogenous sorting leads to PAM, the productively-efficient matching.[10] Instead, the loss of control is related to the manager's decision to commit to $\mathrm{not}$ eliciting reports about workers' types; since wages cannot depend on reports, the manager's ability to extract rents is limited.

# 2 Model

We describe the environment, timing, information, and contracts in the case in which matching is centralized, leaving the description of the delegation alternative to Section 5.

## 2.1 Environment

There is a single profit-maximizing manager, described using female pronouns, and a finite, even number of workers, each described using male pronouns.

[10] Kambhampati et al. (2018) studies delegated matching in an informational environment in which endogenous sorting is not productively efficient.

Workers are indexed by the set $\mathcal{N} := \{1, ..., N\}$ and there at least four workers, $N \geq 4$. Output is produced in teams of two. Within a team, each worker either exerts effort, $e = 1$, or does not, $e = 0$, so that the set of effort levels is $E := \{0, 1\}$. The disutility of effort for every worker is $c > 0$. Each team produces high output, $y = 1$, or low output, $y = 0$, so that the set of output levels is $Y := \{0, 1\}$. The manager is the residual claimant of all output produced by the workers, which may be sold in a competitive market at the normalized price of one per unit. Workers are protected by limited liability and may only receive non-negative wages. Each has an outside option normalized to zero. All parties are risk neutral.

Each worker has a type: high $(H)$, with probability $0 < p < 1$, or low $(L)$, with probability $1-p$. Worker $i$'s type is denoted by $t_i \in T := \{H, L\}$ and is statistically independent of the types of all other workers. The types of workers in a team affect the probability with which the team produces high output: If any worker in a team does not exert effort, his team produces high output with probability zero. But if both workers in a team exert effort, the interior probability with which the team produces high output, $y = 1$, is a symmetric function of its workers' types, $q : T^2 \to (0, 1)$. Let $q_L := q(L, L)$, $q_M := q(L, H) = q(H, L)$, and $q_H := q(H, H)$ denote the probabilities with which a team composed of two lows, one low and one high, and two highs produces high output. Then, $\mathrm{q} := (q_H, q_M, q_L)$ parameterizes the production function.

We make three standard assumptions. First, teams with higher types are, on average, more productive

$$q_H > q_M > q_L.$$

Second, the production technology is strictly supermodular[11]

$$q_H - q_M > q_M - q_L.$$

[11]Equip the product set $T \times T$ with the partial order $\geq$ satisfying $(H, H) \geq (H, L) \geq (L, L)$ and $(H, H) \geq (L, H) \geq (L, L)$. Then $(T \times T, \geq)$ is a complete lattice with $(H, L) \vee (L, H) = (H, H)$ and $(H, L) \wedge (L, H) = (L, L)$ implying that $q$ is a supermodular function.

Third, $c$ is small enough relative to $q_L$ that, in the absence of private information, it is optimal to induce effort by all workers in all teams,

$$q_L > 2c.$$

**Example 1** (Kremer (1993)'s O-Ring Model)**.** *Suppose workers in a team must complete two independent tasks to produce high output, the interior probability with which a high (low) completes his task upon exerting effort is $p_H$ ($p_L$), and highs are more likely than lows to complete their task, $p_H > p_L$. Then, $q_H := p_H * p_H$, $q_M := p_H * p_L$, and $q_L := p_L * p_L$. Simple algebra shows that $q_H > q_M > q_L$ and $q_H - q_M > q_M - q_L$.*

## 2.2 Timing, Information, and Contracts

The initial timing is as follows: first, the manager proposes a contract; second, after learning his own type, each worker accepts or rejects the proposed contract; third, if any worker rejects, no teams are formed and all parties obtain zero utility.[12] If all workers accept the contract, the subsequent timing is as follows:

1. Each worker reports his type to the manager.
2. The manager assigns workers to teams.
3. Each worker learns the type of his assigned teammate.
4. Workers exert effort.
5. The manager observes output and compensates each worker.

A full assignment of workers to teams is a one-to-one function $\nu : \mathcal{N} \to \mathcal{N}$ satisfying (i) $\nu(\nu(i)) = i$ and (ii) $\nu(i) \neq i$, with the interpretation that $\nu(i)$ is

[12] Limited liability and obedience ensure that all workers have an incentive to accept any incentive feasible contract (see Observation 1). Hence, the assumption that the manager does not form teams if any worker does not participate plays no role in our analysis.

worker $i$'s teammate. Condition (i) implies that each worker is his teammate's teammate. Condition (ii) implies that each worker has a teammate. Letting $\mathcal{P}$ denote the set of full assignments, we define a *matching* to be a function associating each vector of reported types with a probability distribution over $\mathcal{P}$,

$$\mu : T^N \to \Delta(\mathcal{P}).$$

A *wage scheme* is a tuple of functions $(w_i)_{i \in \mathcal{N}} := w$, one for each worker $i$, where worker $i$'s wage function maps each full assignment, reported type profile, and output vector to a non-negative wage for that worker,

$$w_i : \mathcal{P} \times T^N \times Y^{N/2} \to \mathbb{R}_+.$$[13]

A *contract* is a matching and a wage scheme.

A comment about the timing is in order. In step 1, the manager elicits all information held among her workers. But, in step 3, due to their close interaction, each worker learns the type of his assigned teammate. Yet, there is no subsequent reporting stage. Contracts are therefore incomplete; mechanisms in which teammates report each other's type are ruled out by assumption.[14] We find this a plausible assumption in environments in which the manager is worried about collusion among her workers or if reporting workers fear retaliation by their co-workers for making undesirable reports.[15] Moreover, maintaining this assumption throughout the paper allows us to make a meaningful comparison between centralized matching and delegated matching. Indeed, any comparison between the two in which delegated matching strictly outperforms centralized matching *requires* the introduction of contractual incompleteness. Otherwise, by the Revelation Principle, delegated matching would appear as a mechanism in the present environment, so that delegation could never result

[13]The co-domain of $w_i$ is restricted to the positive reals because workers are protected by limited liability.

[14]As is well-known, under weak Nash implementation, the manager could obtain full-information profits by punishing workers with zero wages when their reports about the profile of types in their team disagree and compensating them for effort when they agree.

[15]In Section 6, we suggest an approach to endogenize such concerns.

in strictly higher profits.[16]

# 3 The Manager's Problem

To ease the exposition, we restrict attention to contracts that induce effort by all workers in all teams. We discuss relaxing this restriction in Section 6 (a formal analysis can be found in Appendix B).

## 3.1 Full-Information Benchmark

We first consider the full-information benchmark in which both effort and type are contractible. Given any matching, the manager optimally compensates each worker for the cost of effort: each worker receives a wage equal to that cost, $c$, if they exert effort, and zero otherwise. As is well-known, since expected output satisfies strictly increasing differences, the manager then maximizes (minimizes) profits by matching highs with highs (lows) whenever possible. We define the notions of positive and negative assortative matching in our environment before stating this result.

**Definition 1.** *A matching $\mu : T^N \to \Delta(\mathcal{P})$ is a* ***positive assortative matching (PAM)*** *if, for any type profile $\boldsymbol{t} \in T^N$ and any assignment $\nu \in \operatorname{supp} \mu(\boldsymbol{t}) \subseteq \mathcal{P}$, it is never the case that there are two workers $i$ and $j \neq \nu(i)$ for which $t_i \neq t_{\nu(i)}$ and $t_j \neq t_{\nu(j)}$. A matching $\mu : T^N \to \Delta(\mathcal{P})$ is a* ***negative assortative matching (NAM)*** *if, for any type profile $\boldsymbol{t} \in T^N$ and any assignment $\nu \in \operatorname{supp} \mu(\boldsymbol{t}) \subseteq \mathcal{P}$, it is never the case that there are two workers $i$ and $j \neq \nu(i)$ for which $t_i = t_{\nu(i)}$ and $t_j = t_{\nu(j)}$.*

[16]See Poitevin (2000) and Mookherjee (2006) for illuminating discussions. In Section 5.3, we exhibit a matching and wage scheme that can be implemented under delegated matching, but not centralized matching, and explain the fundamental difference between the two environments.

**Theorem** (Becker (1973)). *In the full-information problem, PAM (NAM) maximizes (minimizes) expected profits.*

It is also straightforward to show that the manager attains identical expected profits if type is contractable and effort is not, or vice-versa.[17] Hence, it is the interaction of adverse selection and moral hazard that makes the asymmetric-information problem interesting. We proceed to analyze this case.

### 3.2 Asymmetric-Information Problem

To define the constraints the manager faces in implementing a matching under asymmetric information, we need some notation. Given a matching $\mu$ and truthful reporting by all workers, the interim probability with which worker $i$ matches a high if his type is $t \in T$ is $p_t^{\mu}(i)$;[18] the event that he is assigned to a teammate of type $t' \in T$ is $F_i^{(t,t')}$; and the expected wage he receives given his teammate's type, and when his team produces output $y$ and all workers exert effort, is

$$w_t^{t'}(y,i) := E[w_i(\nu, \mathfrak{t}, y_{(i,\nu(i))}, \mathrm{y}_{-(i,\nu(i))}) | F_i^{(t,t')}].$$

Worker $i$'s expected utility from honesty (reporting her type truthfully) and obedience (always exerting effort), is therefore,

$$\bar{u}_t(i) := \begin{bmatrix} p_t^{\mu}(i) \\ 1 - p_t^{\mu}(i) \end{bmatrix}^T \begin{bmatrix} q(t,H)w_t^H(1,i) + (1-q(t,H))w_t^H(0,i) - c \\ q(t,L)w_t^L(1,i) + (1-q(t,L))w_t^H(0,i) - c \end{bmatrix}.$$

[17] If effort is contractable, but not type, then, as in the full-information case, each worker is compensated for her effort cost and attains zero utility in any team to which she is assigned, independently of her report. Since each worker receives zero utility in any team to which she is assigned, she thus has an incentive to report her type truthfully, so that the manager can implement positive assortative matching.

If, on the other hand, type is contractable, but not effort, then the manager may match highs whenever possible, i.e. implement positive assortative matching. Within each team, since type is observable, the manager can pay each worker the minimal amount required for each worker to exert effort (this, of course, depends on the type of her teammate). As workers are risk neutral, the expected payment to each worker is thus equal to the cost of effort, so that total expected payments are the same as in the full-information case.

[18] Formally,

$$p_t^{\mu}(i) := \sum_{\mathbf{t}_{-i} \in T^{N-1}} Pr(\mathbf{t}_{-i}) \sum_{\nu \in \mathcal{P}} \mu(t, \mathbf{t}_{-i})(\nu) * 1\{t_{\nu(i)} = H\},$$

where $\mu(\mathbf{t})(\nu)$ is the probability of $\nu \in \mathcal{P}$ according to $\mu(\mathbf{t}) \in \Delta(\mathcal{P})$ and $1\{\cdot\}$ is the indicator function.

We now describe the incentive constraints. First, each worker must have a weak incentive to exert effort, given that his teammate exerts effort, in any team to which he is assigned. The obedience constraint for a worker $i \in \mathcal{N}$ of type $t \in T$, given that his teammate has type $t' \in T$, is therefore

$$q(t,t')w_t^{t'}(1,i) - (1-q(t,t'))w_t^{t'}(0,i) - c \geq w_t^{t'}(0,i). \qquad (O_t^{t'}(i))$$

Second, workers must be incentivized to report their types truthfully: each worker's expected payoff under honesty and obedience must exceed the expected payoff he receives from misreporting his type and making an optimal effort decision in each team to which he is assigned. The incentive compatibility constraint for worker $i \in \mathcal{N}$ of type $t \in T$ is therefore,

$$\bar{u}_t(i) \geq \begin{bmatrix} p_{\hat{t}}^{\mu}(i) \\ 1 - p_{\hat{t}}^{\mu}(i) \end{bmatrix}^T \begin{bmatrix} \max\{q(t,H)w_{\hat{t}}^H(1,i) + (1-q(t,H))w_{\hat{t}}^H(0,i) - c, w_{\hat{t}}^H(0,i)\} \\ \max\{q(t,L)w_{\hat{t}}^L(1,i) + (1-q(t,L))w_{\hat{t}}^L(0,i) - c, w_{\hat{t}}^L(0,i)\} \end{bmatrix}, \qquad (IC_t(i))$$

where $\hat{t} \neq t$. Finally, knowing his own type, each worker must be willing to accept the contract proposed by the manager. Equivalently, the utility from honesty and obedience must exceed her outside option of zero. The interim individual rationality constraint for worker $i$ of type $t \in T$ is therefore

$$\bar{u}_t(i) \geq 0. \qquad (IR_t(i))$$

We say that a contract $(\mu, w)$ is incentive feasible if all obedience constraints, all incentive compatibility constraints, and all participation constraints are satisfied. The manager's problem is to chose a contract that maximizes profits– expected output net expected wage payments– subject to the constraint that the contract is incentive feasible.

### 3.3 Redundant Constraints

Before proceeding to the analysis of the optimal contract, we make two observations that reduce the number of constraints. First, the interim individual rationality constraint for each worker is satisfied as long as his obedience constraints are satisfied; any non-negative wages satisfying obedience yield him non-negative expected utility in any team to which he is assigned.

**Observation 1.** *If $O_t^H(i)$ and $O_t^L(i)$ are satisfied, then $IR_t(i)$ is satisfied.*

Second, if the obedience constraints for lows are satisfied, highs have a strict incentive to exert effort after misreporting their type, no matter the type of their assigned teammate.

**Observation 2.** *If $O_L^H(i)$ and $O_L^L(i)$ are satisfied, then $IC_H(i)$ is satisfied as long as,*

$$\bar{u}_H(i) \geq \begin{bmatrix} p_L^\mu(i) \\ 1 - p_L^\mu(i) \end{bmatrix}^T \begin{bmatrix} q_H w_L^H(1,i) + (1-q_H) w_L^H(0,i) - c \\ q_M w_L^L(1,i) + (1-q_M) w_L^L(0,i) - c \end{bmatrix}.$$

In light of Obervation 1 and 2, we henceforth omit participation constraints and “double deviation" constraints by highs when writing the manager’s problem.

# 4 The Optimal Contract

We solve the manager’s problem in three steps. First, we establish that it is without loss of generality to restrict attention to a simple, and tractable, class of matchings and wage schemes. Second, given an arbitrary matching in this class, we identify all optimal wage schemes. Third, given these wage schemes, we identify all profit-maximizing matchings. Our main results, Theorem 1 and Theorem 2, fully characterize the optimal wage schemes and matchings.

## 4.1 Simplifying the Contract Space

The manager’s problem is complex: she faces many constraints and the contract space is large. We greatly simplify her problem by establishing that it is without loss of generality to restrict attention to equal treatment matchings, and independent and anonymous wage schemes.

**Definition 2.** *The matching $\mu$ is an **equal treatment** matching if, for any two workers of the same type, the interim probability with which each is assigned to work with a high is the same,*

$$p_t^\mu(i) = p_t^\mu(j) \text{ for all } i, j \in \mathcal{N} \text{ and } t \in T.$$

**Definition 3.** *The wage scheme $w$ is **independent** if the wage of any worker depends only on his own type, the type of his assigned teammate, and the output his team produces,*

$$w_i(\nu, t_i, \boldsymbol{t}_{-i}, \boldsymbol{y}) = w_i(\nu', t_i, \boldsymbol{t}'_{-i}, \boldsymbol{y}') \text{ when } t_{\nu(i)} = t'_{\nu'(i)} \text{ and } y_{(i,\nu(i))} = y'_{(i,\nu'(i))}.$$

**Definition 4.** *The wage scheme $w$ is **anonymous** if expected wages do not depend on a worker's identity,*

$$w_t^{t'}(y, i) = w_t^{t'}(y, j) \text{ for all } i, j \in \mathcal{N},\ t, t' \in T, \text{ and } y \in Y.$$

**Lemma 1.** *For any incentive feasible contract $(\mu, w)$, there exists another incentive feasible contract, $(\hat{\mu}, \hat{w})$, that attains at least the same expected profits, where $\hat{\mu}$ is an equal treatment matching and $\hat{w}$ is an independent and anonymous wage scheme.*

The restriction to independent wage schemes is without loss of generality because each worker's type and effort decision do not affect the probabilities with which teams other than his own produce high output.[19] Restriction to anonymous wage schemes is without loss of generality because workers

[19] The result is a straightforward application of the Informativeness Principle (Hölmstrom (1979) and Shavell (1979)).

are ex-ante identical and the effort equilibrium the manager implements within teams is symmetric.[20] Finally, restriction to equal treatment matchings is without loss of generality because, for any non-equal treatment matching, the manager attains identical profits using an equal treatment matching that uniformly randomizes over labelings of workers and then applies the original non-equal treatment matching and wages.

Lemma 1 simplifies the manager's problem in two ways. First, any anonymous and independent wage scheme may be summarized by eight wages

$$(w_t^{t'}(y))_{t,t'\in T, y\in Y} \in \mathbb{R}_+^8,$$

where $w_t^{t'}(y)$ denotes the wage given to any worker whose type is $t$, teammate's type is $t'$, and whose team produces output $y$. Second, any equal treatment matching may be summarized by a single parameter $p_H^\mu$, the interim probability with which a worker matches a high upon reporting his type as high.[21] Let $\mathcal{M}$ denote the set of all equal treatment matchings. It will be useful to partition $\mathcal{M}$ into equivalence classes wherein two matchings, $\mu$ and $\mu'$, are in the same equivalence class if $p_H^\mu = p_H^{\mu'}$. Three equivalence classes will be of interest in what follows.

**Definition 5** (Assortative Equal Treatment Matchings)**.** *An equal treatment matching $\mu$ is a **positive assortative matching (PAM)** if*

$$p_H^\mu = \max_{\tilde{\mu}\in\mathcal{M}} \; p_H^{\tilde{\mu}},$$

[20]In contrast to the setting of Winter (2004), who shows that non-anonymous schemes may be useful when the manager wants to implement a *unique* Nash Equilibrium, we require only that effort by two workers in a team is *a* Nash Equilibrium. Indeed, given the production technology and the assumption of limited liability, requiring uniqueness of the high effort equilibrium is impossible: low effort by any worker in a team dooms a project to failure, and effort is costly, so that there is always a Nash Equilibrium in which both workers do not exert effort.

[21]Let $p_L^\mu$ denote the interim probability with which a worker matches a high upon reporting her type as low. In every matching, the number of highs matched to lows must equal the number of lows matched to highs. Hence, the ex-ante probability with which a worker is a high and matches a low must equal the ex-ante probability with which a worker is a low and matches a high, i.e. $p(1-p_H^\mu) = (1-p)p_L^\mu$. Therefore, $p_L^\mu = \frac{p}{1-p}(1-p_H^\mu)$ is determined by $p_H^\mu$.

*a **negative assortative matching (NAM)** if,*

$$p_H^\mu = \min_{\tilde{\mu}\in\mathcal{M}} \; p_H^{\tilde{\mu}},$$

*and a **random matching (RM)** if,*

$$p_H^\mu = p.$$

It is easy to show that any PAM (NAM) is a positive (negative) assortative matching in the sense of Definition 1.

### 4.2 The Minimization Problem

Given Lemma 1, the manager's minimization problem is to choose an anonymous, independent wage scheme, described by the tuple $(w_t^{t'}(y))_{t,t'\in T, y\in Y} \in \mathbb{R}_+^8$, to minimize expected wage payments subject to the incentive constraints,

$$[IC_H] \;\; \bar{u}_H \geq \begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_L^H(1) + (1-q_H) w_L^H(0) - c \\ q_M w_L^L(1) + (1-q_M) w_L^L(0) - c \end{bmatrix}$$

$$[IC_L] \;\; \bar{u}_L \geq \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} \max\{q_M w_H^H(1) + (1-q_M) w_H^H(0) - c, w_H^L(0)\} \\ \max\{q_L w_H^L(1) + (1-q_L) w_H^L(0) - c, w_H^L(0)\} \end{bmatrix}$$

$$[O_H^H] \;\; q_H w_H^H(1) + (1-q_H) w_H^H(0) - c \geq w_H^H(0)$$

$$[O_H^L] \;\; q_M w_H^L(1) + (1-q_M) w_H^L(0) - c \geq w_H^H(0)$$

$$[O_L^H] \;\; q_M w_L^H(1) + (1-q_H) w_L^H(0) - c \geq w_L^H(0)$$

$$[O_L^L] \;\; q_L w_L^L(1) + (1-q_L) w_L^L(0) - c \geq w_L^L(0).$$

The first issue the manager must resolve is standard. In order to incentivize a low to exert effort, she must pay him a high enough wage for producing high output relative to what she pays him for producing low output. But, fixing teammate type, highs are more likely than lows to produce high output. Hence, highs must always be paid a strictly positive information rent in order to prevent them from misreporting their type. We show that, in any optimal wage scheme, the incentive compatibility constraint for highs and the obedience constraints for lows bind, and that highs always receive zero wages when they produce low output.

**Lemma 2.** *In any optimal wage scheme, $w_H^H(0) = 0$, $w_H^L(0) = 0$ and $IC_H$, $O_L^H$, and $O_L^L$ bind.*

The second issue the manager must resolve is novel to our environment. In addition to the downward incentive compatibility constraint, $IC_H$, the upward incentive compatibility constraint, $IC_L$, may also bind. In particular, $IC_L$ binds when the implemented matching exhibits positive (negative) assortativity and the production technology is strictly log submodular (supermodular).

**Definition 6.** *$\mu$ exhibits **positive assortativity** if $p_H^\mu > p$ and **negative assortativity** if $p_H^\mu < p$.*[22]

**Definition 7.** *The production technology is **log supermodular** if $\frac{q_H}{q_M} \geq \frac{q_M}{q_L}$ and **log submodular** if $\frac{q_H}{q_M} \leq \frac{q_M}{q_L}$.*

**Example 2** (Kremer (1993)'s O-Ring Model Continued)**.** *In the O-Ring model, output probabilities are multiplicative: $q_H := p_H * p_H$, $q_M := p_H * p_L$, and $q_L := p_L * p_L$, where $0 < p_L < p_H < 1$. Hence, the production technology is both log supermodular and log submodular:*

$$\frac{q_H}{q_M} = \frac{p_H * p_H}{p_H * p_L} = \frac{p_H}{p_L} = \frac{p_H * p_L}{p_L * p_L} = \frac{q_M}{q_L}.$$

*The O-Ring model is therefore a "knife-edge" case.*

[22] By the observation made in Footnote 21, positive assortativity is equivalent to $p_H^\mu > p_L^\mu$ and negative assortativity is equivalent to $p_H^\mu < p_L^\mu$.

In these cases, the manager optimally offers two wage schemes: one targeted to lows in which low output is sometimes rewarded with positive wages, and one targeted to highs in which low output is never rewarded with positive wages.

**Theorem 1** (Optimal Wages)**.**

- *If $\mu$ exhibits positive (negative) assortativity and the production technology is strictly log submodular (supermodular), then either $w_L^H(0) > 0$ or $w_L^L(0) > 0$ and both types of workers receive a strictly positive information rent: $\bar{u}_L > 0$ and $\bar{u}_H > 0$.*
- *If $\mu$ exhibits positive (negative) assortativity and the production technology is log supermodular (submodular), then $w_L^H(0) = w_L^L(0) = 0$ and only highs receive a strictly positive information rent: $\bar{u}_L = 0$ and $\bar{u}_H > 0$.*

The result is driven by a feasibility issue that arises when the manager attempts to hold lows to their reservation utility. If $\mu$ exhibits positive (negative) assortativity and the production technology is log supermodular (submodular), it is incentive feasible, and also optimal, to do so. But if $\mu$ exhibits positive (negative) assortativity and the production technology is strictly log submodular (supermodular), there do not exist wages holding lows to their reservation utility that satisfy both $IC_L$ and $IC_H$. The intuition for the result is as given in the introduction: under strict log submodularity, highs strictly prefer to match a low after misreporting their type and this is more likely under a matching exhibiting positive assortativity. Dissuading such deviations requires increasing wages to highs when they report their type truthfully. But increasing these wages makes it profitable for lows to masquerade as highs. Consequently, the upward incentive compatibility constraint binds at the optimal wage scheme and lows must be paid a strictly positive information rent. We outline this issue, leaving the full proof of Theorem 1, along with closed-form expressions of the optimal wages, to the Appendix.

#### 4.2.1 The Problem of Disassortative Incentives

By Lemma 2, we know that in any optimal wage scheme both obedience constraints for lows, $O_L^H$ and $O_L^L$, bind:

$$q_M w_L^H(1) + (1 - q_M) w_L^H(0) - c = w_L^H(0),$$

and,

$$q_L w_L^L(1) + (1 - q_L) w_L^L(0) - c = w_L^L(0).$$

Hence, the utility lows receive at the optimal wage scheme is entirely determined by the wages they receive upon producing low output,

$$\bar{u}_L = \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} q_M w_L^H(1) + (1 - q_M) w_L^H(0) - c \\ q_L w_L^L(1) + (1 - q_L) w_L^L(0) - c \end{bmatrix} = \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} w_L^H(0) \\ w_L^L(0) \end{bmatrix}.$$

A natural approach to solving the manager's problem, then, is to minimize wage payments to lows, i.e. set $w_L^H(0) = w_L^L(0) = 0$, so that lows do not obtain rent in excess of their reservation utility, $\bar{u}_L = 0$. Indeed, it is straightforward to show that it is optimal to do so when $\mu$ exhibits positive (negative) assortativity and the production technology is log supermodular (submodular).

What goes wrong in the other cases? By Lemma 2, $w_H^H(0) = w_H^L(0) = 0$ in any optimal wage scheme. Hence, when $\bar{u}_L = 0$, the incentive compatibility constraint for lows simplifies to,

$$\bar{u}_L = 0 \geq \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} \max\{q_M w_H^H(1) - c, 0\} \\ \max\{q_L w_H^L(1) - c, 0\} \end{bmatrix}. \tag{$IC_L$}$$

We see that if $w_H^H(1) > \frac{c}{q_M}$ ($w_H^L(1) > \frac{c}{q_L}$), then regardless of the value of $w_H^L(1)$ ($w_H^H(1)$), a low could misreport his type and obtain strictly positive utility by exerting effort only when matched with a high (low).

It turns out, however, that there do not exist wages $w_H^H(1) \leq \frac{c}{q_M}$ and $w_H^L(1) \leq \frac{c}{q_L}$ that also satisfy the incentive compatibility constraint for highs, $IC_H$, if $\mu$ exhibits positive (negative) assortativity and the production technology is strictly log submodular (supermodular). In other words, in these cases, it is impossible to satisfy both $IC_L$ and $IC_H$ while keeping lows to their reservation utility. To see why, notice that the payoff to highs from truth-telling is given by,

$$\bar{u}_H = \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix},$$

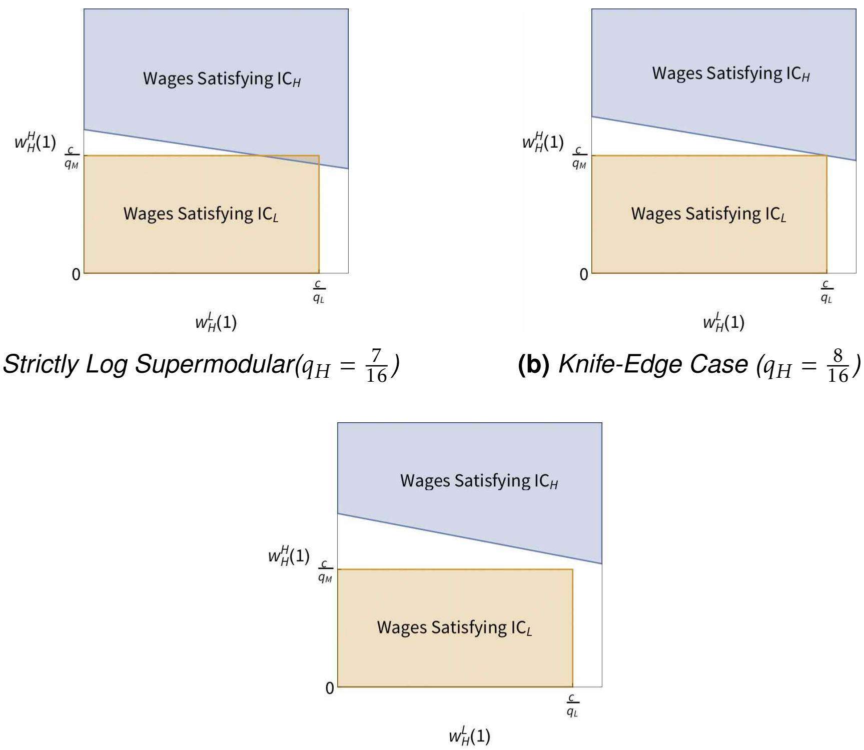
while the payoff to deviating is given by,

$$\begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_L^H(1) + (1 - q_H) w_L^L(0) - c \\ q_M w_L^L(1) + (1 - q_M) w_L^L(0) - c \end{bmatrix} = \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c - c \\ \frac{q_M}{q_L} c - c \end{bmatrix},$$

where the equality follows by setting $w_L^H(0) = w_L^L(0) = 0$ and noticing that $w_L^H(1) = \frac{c}{q_M}$ and $w_L^L(1) = \frac{c}{q_L}$ by the binding obedience constraints, $O_L^H$ and $O_L^L$. Eliminating effort costs from both sides of the equation, we obtain a simple expression for $IC_H$ ,

$$\begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) \\ q_M w_H^L(1) \end{bmatrix} \geq \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c \\ \frac{q_M}{q_L} c \end{bmatrix}. \tag{$IC_H$}$$

Inspecting $IC_H$, we notice that if $\mu$ exhibits positive assortativity, so that $p_H^\mu > p$, and hence $p_H^\mu > p_L^\mu$, then a high is more likely to match a low after misreporting his type than after reporting truthfully. But the production technology is strictly submodular, so that $\frac{q_H}{q_M} < \frac{q_M}{q_L}$, then a high strictly prefers to match a low after misreporting. So, in order to prevent the high from deviating, the manager must give him a high enough wage, $w_H^H(1)$, when he reports truthfully, matches a high, and produces high output. Unfortunately for the manager, if $w_H^L(1) \leq \frac{c}{q_M}$, then it must be that $w_H^H(1) > \frac{c}{q_M}$ in order to satisfy $IC_H$. Hence, any wage scheme satisfying $IC_H$ violates $IC_L$. See Figure 1 for an illustration.[23]




**(a)** *Strictly Log Supermodular*($q_H = \frac{7}{16}$) **(b)** *Knife-Edge Case* ($q_H = \frac{8}{16}$)

**(c)** *Strictly Log Supermodular* ($q_H = \frac{9}{16}$)

**Figure 1.** *Positive Assortative Matching:* $N = 4$, $q_M = \frac{4}{16}$ , $q_L = \frac{2}{16}$, $p = \frac{1}{2}$, $\bar{u}_L = 0$.

When it is infeasible to hold lows to their reservation utility, both incentive compatibility constraints bind at the optimal wage scheme, and both types receive information rents. Indeed, giving higher wages to highs to prevent downward deviations entails giving higher wages to lows to prevent upward deviations, and vice-versa. Luckily, the manager can resolve this cyclicality by giving lows “low-powered" incentives, simultaneously increasing their wages upon producing low and high output, and highs “high-powered" incentives, only increasing their wages upon producing high output. These schemes come at a cost, however, and may induce the manager to distort the matching in order to save on expected wage payments. We turn to this issue next.

[23]A similar issue occurs when the technology is strictly supermodular and $\mu$ exhibits negative assortativity.

### 4.3 The Maximization Problem

Maximizing (minimizing) the assortativity of the implemented matching, $p_H^\mu$, clearly maximizes (minimizes) expected output; indeed, any such matching is a positive (negative) assortative matching. Absent incentive costs, PAM is therefore optimal, as pointed out in Section 3.1. It turns out, however, that when the production technology is strictly log submodular, optimal expected wage payments under asymmetric information are $\text{strictly increasing}$ in $p_H^\mu$. A non-trivial rent-efficiency tradeoff arises. Theorem 2, our main result, fully characterizes the optimal matchings in terms of the parameters of the model.

**Theorem 2** (Optimal Matching)**.** *If the production technology is log supermodular, then PAM is the unique optimal matching. If the production technology is strictly log submodular, however, there exist two cutoff values on the cost of effort,* $0 < \underline{c} < \bar{c}$*, such that,*

1. *PAM is the unique optimal matching if and only if* $c < \underline{c}$*;*
2. *RM is the unique optimal matching if and only if* $\underline{c} < c < \bar{c}$*; and,*
3. *NAM is the unique optimal matching if and only if* $c > \bar{c}$*.*

*Further, under strict log submodularity, RM becomes optimal for all effort costs as highs vanish from the population and NAM becomes optimal for all effort costs as lows vanish from the population.*

A number of aspects of Theorem 2 are worthy of attention. First, log supermodularity is a sufficient condition for PAM, coinciding with the condition for PAM found by Smith (2006). Second, under strictly log submodularity, PAM is suboptimal in a variety of circumstances–both when the effort cost is sufficiently high and when the prior probability of highs is close enough to zero or one. Third, despite the complex matchings the manager has available to her, the optimal matchings are simple. To implement PAM (NAM), the manager need only match highs with highs (lows), uniformly randomizing which worker is matched with a low (high) in the case of an odd number of highs. To implement RM, the manager need only commit to a full assignment and disregard any reports she receives.[24] Fourth, the optimal matchings we identify are unique up to their assortativity properties, outside the special cases in which $c = \bar{c}$ and $c = \underline{c}$. Fifth, our proof provides a full quantitative characterization of the optimal matchings, with exact values of the cutoffs $\underline{c}$ and $\bar{c}$ as determined by the prior $p$ and the production parameters $\mathrm{q}$.[25] See Figure 2 for an illustration.

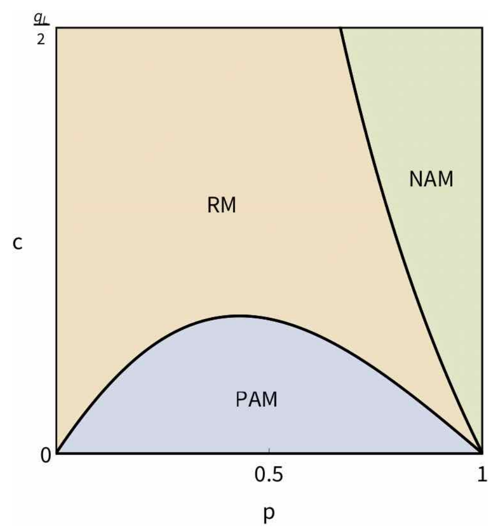



**Figure 2.** *Optimal Matchings:* $q_H = \frac{7}{16}$, $q_M = \frac{4}{16}$, $q_L = \frac{2}{16}$.

#### 4.3.1 Intuition

We discuss the driving forces behind each result. From Theorem 1, we know that, under log supermodularity, any matching exhibiting positive assortativity

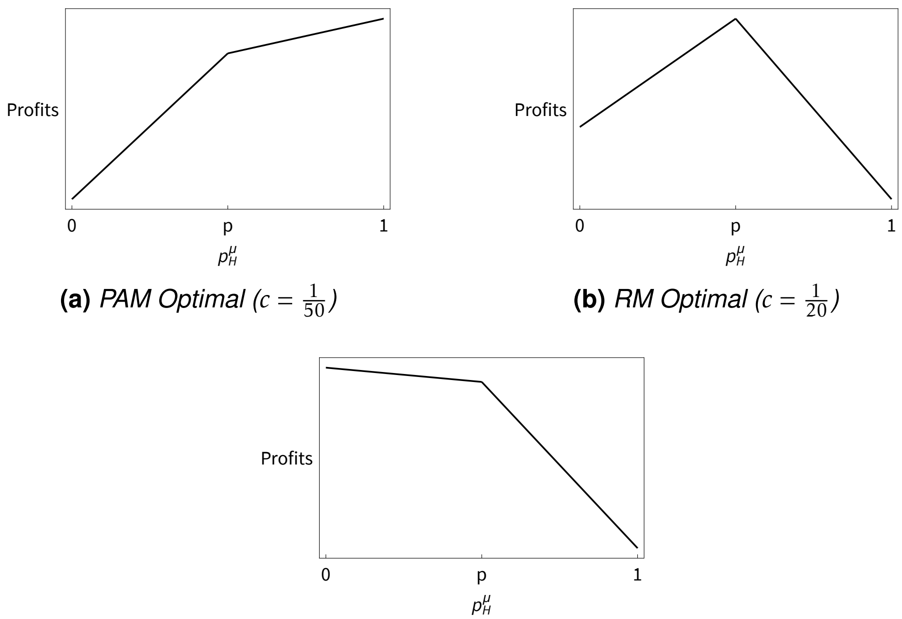
[24]Any matching which does not depend on reports, i.e. any constant function $\mu : T^N \to \Delta(\mathcal{P})$, is a random matching according to our definition. However, there are non-constant random matchings.

[25]The formulas are determined by the likelihood ratio of lows to highs, $\frac{1-p}{p}$, complementarity in the production function in terms of levels, $q_H + q_L - 2q_M$, and in terms of ratios, $\frac{q_M}{q_L} - \frac{q_H}{q_M}$. The exact expression is given in the proof of the Theorem in the Appendix.

minimizes expected wage payments to lows– lows are held to their reservation utility. It turns out that, in addition, expected wage payments to highs are minimized under PAM. Indeed, after misreporting his type, a high is more likely to match a low, obtaining a utility gain proportional to $\frac{q_M}{q_L}$ instead of $\frac{q_H}{q_M}$. And, under log supermodularity, $\frac{q_M}{q_L} \leq \frac{q_H}{q_M}$. Hence, to deter downward deviations, expected wage payments to highs must decrease as the implemented matching becomes more assortative. It follows that the output-efficient matching, PAM, is optimal.

From Theorem 1, we also know, however, that in the case of strict log submodularity, if the manager implements PAM, then lows must obtain wage payments in excess of those at which they attain their reservation utility. Applying the opposite logic of what we described previously, as the implemented matching becomes more assortative, wage payments to highs must increase to deter downward deviations. It follows that total expected wage payments are increasing in assortativity. Hence, if the reduction in incentive costs accrued from implementing an inefficient matching is sufficient to compensate the manager for the loss in expected output, she optimally distorts PAM. This happens precisely when $c > \underline{c}$.

When the manager distorts PAM, why does she sometimes choose RM and other times NAM? As the assortativity of the implemented matching decreases, the slope of optimal expected wage payments decreases precisely at the point at which the implemented matching becomes random, $p_H^\mu = p$. This happens because, below this point, the manager no longer saves money by reducing wage payments to lows; indeed, lows are held to their reservation utility at the optimal wage scheme for any matching exhibiting negative assortativity. It follows that any reduction in incentive costs when reducing $p_H^\mu$ in the region $[0, p]$ owes to decreasing expected wage payments to highs only. Hence, there exists an interval of effort costs, $(\underline{c}, \bar{c})$, at which it is only optimal to distort the matching until the point at which $p_H^\mu = p$, so that only RM is optimal. If, however, $c > \bar{c}$, then the gain of reducing payments to highs by reducing the assortativity of the matching becomes so large that it is optimal to implement NAM. Figure 3 depicts expected profits as a function of $p_H^\mu$ in all three cases.




**(a)** *PAM Optimal* ($c = \frac{1}{50}$)

**(b)** *RM Optimal* ($c = \frac{1}{20}$)

**(c)** *NAM Optimal* ($c = \frac{1}{5}$)

**Figure 3.** *Profits under strict log submodularity:* $q_H = \frac{7}{16}$, $q_M = \frac{4}{16}$, $q_L = \frac{2}{16}$, $p = \frac{1}{2}$.

We now explain why PAM becomes suboptimal for all effort costs as the probability of highs becomes extreme. As $p$ approaches one or zero, it is clear that the difference in expected output between PAM and RM (or any other matching) disappears linearly. But it is less clear that the difference in expected wage payments under PAM and RM disappears sublinearly. This happens for the following reason. In order to implement PAM, the manager must give strictly positive utility to lows to deter upward deviations. This implies that she must also give highs an additional utility payment to deter downward deviations, relative to what she gives them when implementing RM. Hence, as $p$ approaches one (zero), so that lows (highs) vanish in the population, the mere existence of lows (highs) means that the manager must pay highs (lows) an additional utility payment along the entire sequence. Expected wage payments, therefore, vanish at a slower rate than the actual population of lows (highs).

A final result remains to be explained: why does NAM come to dominate RM as $p$ approaches one, but not as $p$ approaches zero? Under both NAM and RM, wage payments to lows are minimized. NAM and RM differ, however, in terms of expected wage payments to highs–the less assortative the matching, the lower these payments. When $p$ is small, so that highs are rare, the difference between RM and NAM in terms of wage payments to highs is outweighed by the productivity gain of RM relative to NAM. But as $p$ grows large, so that highs are common, productivity gains matter more than wage payments.

# 5 Delegated Matching

As we have seen, it may be costly to implement the efficient matching in a centralized workplace in which a manager asks each worker to report his type, and uses these reports to assign workers to teams. But what if, instead, the manager did not ask for reports, and simply allowed workers to sort themselves?

A familiar tradeoff arises. On one hand, such an arrangement entails a loss of control for the manager: She can no longer tailor wages to reports. But on the other hand, the manager can exploit local information: It is reasonable to think that workers possess superior information about one another's characteristics and that they might use this information to sort efficiently.

We formalize this tradeoff by considering an environment in which there is no reporting stage, but in which, during the process of finding a teammate, workers commonly learn the true type profile. Using this knowledge, workers then form self-enforcing teams.

## 5.1 Timing, Information, and Contracts

The environment is the same as in Section 2.1. Moreover, the initial timing is the same as in the case of centralized matching: First, the manager proposes a contract; second, after learning his own type, each worker accepts or rejects the proposed contract; third, if any worker rejects, no teams are formed and all parties obtain zero utility. In contrast to the centralization environment, however, there is no reporting stage. Instead, workers learn one another's types and use this information to form teams. Formally, the timing after contracts have been signed is as follows:

1. Workers commonly learn each other's types.
2. Workers form teams.
3. Workers exert effort.
4. The manager observes output and compensates each worker.

A delegation contract is thus a wage scheme $w := (w_i)_{i\in\mathcal{N}}$ in which each wage function $w_i$ depends only on a worker's identity, the realized assignment, and observed output, i.e. for all $i$,

$$w_i : \mathcal{P} \times Y^{N/2} \to \mathbb{R}_+.$$

As before, we focus on contracts that induce effort by every worker in every team, leaving a discussion of this restriction to Section 6.

### 5.2 Manager's Problem

Though not formally part of a contract, we may think of the manager as choosing a matching function, $\mu : T^N \to \Delta(\mathcal{P})$, in addition to a wage scheme $w$. In contrast to the centralization environment, however, to implement a matching $\mu$, we require that any assignment realized with positive probability under $\mu$ must be self-enforcing given $w$ in the sense that it is in the core. We call any matching satisfying this property stable.

**Definition 8.** *Given a wage scheme $w$, an assignment $\nu \in \mathcal{P}$ is in the* ***core*** *if, for any worker $i \in \mathcal{N}$ and any worker that is not $i$'s teammate, $j \neq \nu(i)$,*

1. *$i$ and $\nu(i)$ exerting effort is a Nash Equilibrium; and*

2. *$i$ and $j$ cannot form a deviating team and select a Nash Equilibrium making each strictly better off.*

*A matching $\mu$ is **stable** with respect to $w$ if for any realized type profile $\mathbf{t} \in T^N$ and any assignment $\nu \in \text{supp } \mu(\mathbf{t})$, $\nu$ is in the core given $w$.*

The manager's delegation problem is to choose a matching and a wage scheme, $(\mu, w)$, to maximize profits, subject to the constraint that $\mu$ is stable given $w$.

### 5.3 An Implementation Unattainable Under Centralized Matching

Before proceeding to identify the optimal delegation contract, we point out that some combinations of matchings and wage schemes that can be implemented under delegation cannot be implemented under centralized matching.[26] For instance, suppose the manager pays each worker a wage of $\frac{c}{q_L}$ if his team produces high output and $0$ if his team produces zero output. Then, PAM is stable respect to this wage scheme. To see this, note that exerting effort is an equilibrium even in teams composed of two lows; in such a team, the payoff from exerting effort when one's teammate does is $q_L \frac{c}{q_L} - c = 0$, the payoff of not exerting effort. In addition, there are no strictly profitable deviating teams; two distinct highs are never matched with a low under PAM. Finally, a low can never form a strictly profitable team with a high, as any such high would be made weakly worse off.

Notice, however, that PAM cannot be implemented using this wage scheme under centralized matching. If the manager were to ask workers to report their type in order to implement PAM, lows would never have an incentive to report their type truthfully. Indeed, as wages do not depend on type and exerting effort is an equilibrium in any team, every worker strictly prefers to match a high. Hence, under PAM, lows have a strict incentive to masquerade as highs. The

[26]The opposite holds as well; Lemma 3, below, shows that under delegated matching anonymous and independent wage schemes can only implement PAM.

fundamental difference between delegated matching and centralized assignment is thus that, under delegated matching, the manager commits not to ask workers report their type and instead empowers workers to use the information they have about one another to sort themselves into teams.

### 5.4 Optimal Delegation Contract

We now show that the contract described in Section 5.3 is actually optimal. As before, say that a delegation contract is anonymous if it does not depend on the worker's identity and independent if the worker's wage depends only on output produced in her own team. An anonymous, independent wage scheme may therefore be represented by a pair $(w(1), w(0)) \in \mathbb{R}_+^2$, specifying a non-negative wage following each output level. It turns out that the only stable matchings the manager can implement with respect to an anonymous and independent wage scheme are PAM.

**Lemma 3.** *If $w$ is an anonymous and independent wage scheme, then $\mu$ is stable with respect to $w$ only if $\mu$ is a PAM.*

*Proof.* Suppose that $w$ is an anonymous and independent wage scheme and that $\mu$ is not a PAM. Then, there exists some type profile $\mathbf{t} \in T^N$, assignment $\nu \in \operatorname{supp} \ \mu(\mathbf{t})$, and workers $i$ and $j \neq \nu(i)$ such that $t_i \neq t_{\nu(i)}$ and $t_j \neq t_{\nu(j)}$. Without loss of generality, suppose that $t_i = t_j = H$, so that $t_{\nu(i)} = t_{\nu(j)} = L$.

To show that $\mu$ is not stable with respect to $w$, we show that $\nu$ is not in the core. Towards contradiction, suppose that $\nu$ is in the core. Then, in each team formed under $\nu$, it must be that effort by both workers in teams $(i, \nu(i))$ and $(j, \nu(j))$ is a Nash Equilibrium. This is the case if and only if,

$$q_M w(1) + (1 - q_M) w(0) - c \geq w(0) \iff w(1) - w(0) \geq \frac{c}{q_M} > 0.$$

But, since $q_H > q_M$, this implies that,

$$q_H w(1) + (1 - q_H) w(0) - c > q_M w(1) + (1 - q_M) w(0) - c.$$

It follows that $i$ and $j$ obtain a strictly higher payoff by forming a deviating team and exerting effort. Hence, $v$ cannot be in the core, our desired contradiction.

□

But PAM is efficient! Hence, the anonymous and independent wage scheme that minimizes expected wage payments yields the manager at least the same profits as any other delegation contract.

**Lemma 4.** $(w(1), w(0)) = (\frac{c}{q_L}, 0)$ *is an optimal delegation contract.*

*Proof.* As observed previously, PAM is stable with respect to $(\frac{c}{q_L}, 0)$. Moreover, due to limited liability, wages must always exceed zero after the production of low output. Hence, to ensure that effort is optimal when all workers are lows, a realization that occurs with positive probability, expected wages following high output must be at least $\frac{c}{q_L}$. The result then follows because PAM is output-efficient and wage payments are minimized following the production of low and high output. □

Two comments about the optimal contract are in order. First, since wages depend only on the output produced in a worker's own team, each worker prefers to match a high. Due to common knowledge of types, a high is only willing to match a low if he has no better option, i.e. all other highs are matched with highs already. But these forces lead to PAM, the efficient matching. This is the local information benefit of delegation. Second, as wages cannot be tailored to reported types, both highs and lows receive strictly positive expected utility at the optimal contract; lows receive strictly positive expected utility in the case in which they match a high, and highs receive strictly positive expected

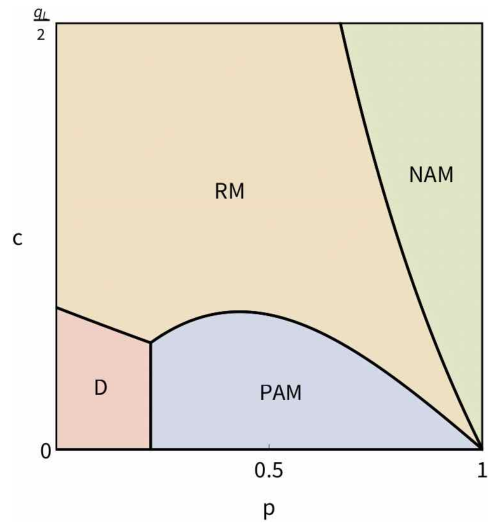



**Figure 4.** *Optimality of Delegation ("D"):* $N \to \infty$, $q_H = \frac{7}{16}$, $q_M = \frac{4}{16}$ , $q_L = \frac{2}{16}$.

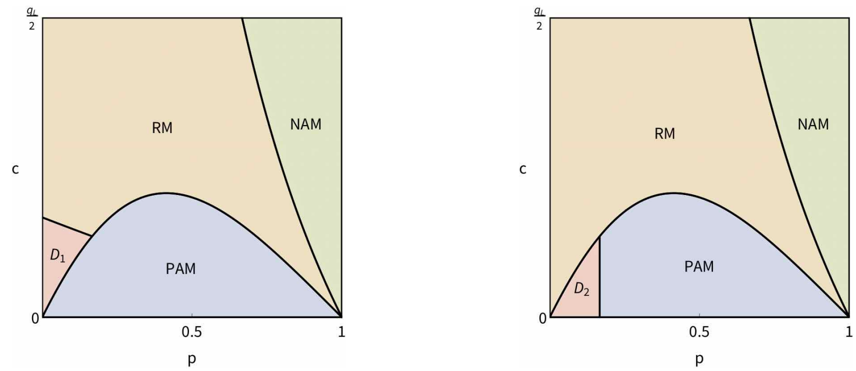
utility no matter their teammate. This is the loss of control cost of delegation. Delegation is optimal if and only if the local information benefit outweighs the loss of control cost.

## 5.5 Optimality of Delegation

Theorem 3, our final result, identifies sufficient conditions under which delegation outperforms centralized assignment. In its statement, and in the subsequent discussion, we abuse notation and let PAM, RM, and NAM denote the scenarios in which the manager chooses centralized assignment and implements each of these matchings at minimal expected cost.

**Theorem 3** (Optimality of Delegation)**.** *If the production technology is log supermodular, then delegation is strictly suboptimal. If the production technology is strictly log submodular, however, then in the limit as the number of workers grows large, there exists an effort cost $\tilde{c} \in (0, 1)$ and a prior $\tilde{p} \in (0, 1)$ such that delegation is optimal if and only if $c \leq \tilde{c}$ and $p \leq \tilde{p}$.*

See Figure 4 for an illustration and the proof of Theorem 3 for exact expressions of $\tilde{c}$ and $\tilde{p}$.



**(a)** $D_1$*: Delegation Optimal when RM Optimal.*

**(b)** $D_2$*: Delegation Optimal when PAM Optimal.*

**Figure 5.** *Decomposition of Optimal Delegation Region.*

To understand the role of the conditions that production is strictly submodular and $N$ is large, it is instructive to recall the logic behind Theorem 2. Under strict log submodularity, a rent-efficiency tradeoff arises because a high worker, counterintuitively, would rather match a low than a high after misreporting his type. But as the number of workers in the firm increases, the probability with which a worker is assigned to work with a teammate with the same reported type approaches one. Hence, the probability with which a deviating high matches a low after misreporting his type approaches one. The two conditions together ensure that incentive costs under PAM exceed those under the optimal delegation contract; the first ensures that there is a rent-efficiency tradeoff, while the second ensures that this tradeoff is large enough. Notice that if the joint distribution over type profiles was such that there were always an $\text{even}$ number of highs, then no condition on $N$ would be required.

One additional condition, that the probability of highs, $p$, is small enough, is required to ensure that delegation outperforms PAM. This ensures that the loss of control under delegation is sufficiently small. In particular, as the probability of highs becomes small, under delegation, the rent paid to workers on average shrinks to zero because a low obtains zero expected utility when matched with another low. However, under centralized assignment, the mere existence of highs makes incentivizing truth-telling by lows costly. Hence, when $p$ is small enough, delegation results in smaller expected wage payments than PAM. See Figure 5b for a depiction of the region in which PAM is optimal under centralization, but delegation outperforms PAM.

The final condition, that the cost of effort $c$ is small enough, is the most interesting. Under centralized assignment, when talent is scarce and the cost of effort is high enough, RM dominates PAM; the manager prefers to introduce productive inefficiency to reduce wages. We show, however, that there is a region in which RM is optimal under centralization, but delegation is optimal overall (see Figure 5a). This result owes to the local information benefit of delegation. Rather than commit to a distorted matching and elicit reports, the manager would rather allow workers to sort themselves because workers, utilizing common knowledge of each other's types, will sort efficiently. In other words, if a manager contemplates distorting the efficient matching in order to pay lower wages, there may be another option: write a "pay-for-performance" contract that does not depend on non-verifiable reports, and simply allow workers to sort themselves.

We conclude our analysis of delegation by remarking that the tradeoff between centralized matching and delegation is interesting $\text{only}$ in the case in which production is supermodular. If the production technology were to be submodular, then our analysis of the cost-minimizing wage scheme given NAM indicates that expected wage payments are strictly lower than under delegated matching. Moreover, endogenous sorting leads to PAM, which is inefficient given a submodular production technology. Hence, centralized matching dominates delegation both in terms of extracting rent and productive efficiency.

# 6 Discussion

Our analysis identifies a new channel by which asymmetric information, in the form of hidden effort and private information, distorts PAM. For strictly log submodular production technologies, more productive workers benefit more from matching less productive workers after misreporting their type, leading to the problem of disassortative incentives (Theorem 1). Disassortative incentives give rise to a novel rent-efficiency tradeoff: if the cost of effort is sufficiently high, then either RM or NAM is optimal. Furthermore, RM becomes optimal for all effort costs as talent becomes scarce and NAM becomes optimal for all effort costs as talent becomes abundant (Theorem 2). We investigate the implications of these results for the optimal management of teams inside the firm, and find conditions under which delegating the sorting problem to workers outperforms centralized assignment (Theorem 3). Together, our results rationalize recent evidence of non-assortative matching inside of firms.

We conclude by discussing two important assumptions maintained in our analysis. First, we have assumed, but not verified, that the manager finds it optimal to implement effort by all workers. When talent is scarce, so that the prior probability of highs is small, implementing effort by all workers is clearly optimal. Hence, the distortion of PAM we identify is robust to alternative effort implementations. When talent is abundant, however, it is not immediate that inducing effort by all workers is optimal. Hence, it is no longer clear that distorting PAM is optimal. In Appendix B, in the case in which talent is abundant and the firm is large, we provide conditions on the production function, stronger than those of strict log submodularity, under which NAM and effort by all workers outperforms PAM and any effort implementation. It follows that distorting PAM is globally optimal in these cases.

Second, we do not investigate mechanisms that provide information to the manager after teams have been formed. While relaxing this assumption is theoretically interesting, we find this restriction plausible in environments in which peer evaluation is ineffective at generating reliable reports. As previously mentioned, two explanations for why this may be the case are that peer reports may be subject to collusion and reporting parties may fear retaliation by their co-workers (see Che y Yoo (2001), who make a similar non-contractability assumption, for a discussion of the former point and Chassang y Zehnder (2019) for recent work related to the latter point). An interesting, though challenging, direction for future research would be to study whether the distortion we identify holds in a dynamic contracting environment in which information about one's teammate arrives over time and the manager demands contracts to be robust to collusion and/or that individual reports cannot be identified by the manager's chosen matching.

## References


Adhvaryu, A., Bassi, V., Nyshadham, A., y Tamayo, J. A. (2019). No line left behind: Assortative matching inside the firm. SSRN Electronic Journal.

Aghion, P. y Tirole, J. (1997). Formal and real authority in organizations. Journal of Political Economy, 105(1), 1–29.

Anderson, A. y Smith, L. (2010). Dynamic matching and evolving reputations. The Review of Economic Studies, 77(1), 3–29.

Becker, G. S. (1973). A theory of marriage: Part i. Journal of Political Economy, 81(4), 813–846.

Chade, H. y Eeckhout, J. (2018). Matching information. Theoretical Economics, 13(1), 377–414.

Chassang, S. y Zehnder, C. (2019). Secure survey design in organizations: Theory and experiments. Working Paper 25918, National Bureau of Economic Research.

Che, Y.-K. y Yoo, S.-W. (2001). Optimal incentives for teams. American Economic Review, 91(3), 525–541.

Chiappori, P.-A. y Reny, P. J. (2016). Matching to share risk. Theoretical Economics, 11(1), 227–251.

Damiano, E. y Li, H. (2007). Price discrimination and efficient matching. Economic Theory, 30(2), 243–263.

Dessein, W. (2002). Authority and communication in organizations. Review of Economic Studies, 69(4), 811–838.

Eeckhout, J. y Kircher, P. (2010). Sorting and decentralized price competition. *Econometrica*, *78*(2), 539–574.

Franco, A. M., Mitchell, M., y Vereshchagina, G. (2011). Incentives and the structure of teams. *Journal of Economic Theory*, *146*(6), 2307 – 2332.

Gomes, R. y Pavan, A. (2016). Many-to-many matching and price discrimination. *Theoretical Economics*, *11*(3), 1005–1052.

Hölmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, *10*(1), 74–91.

Hölmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*(2), 324–340.

Innes, R. D. (1990). Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, *52*(1), 45 – 67.

Johnson, T. (2013). Matching through position auctions. *Journal of Economic Theory*, *148*(4), 1700 – 1713.

Kambhampati, A., Segura-Rodriguez, C., y Shao, P. (2018). Matching to produce information. *SSRN Electronic Journal*.

Kaya, A. y Vereshchagina, G. (2014). Partnerships versus corporations: Moral hazard, sorting, and ownership structure. *American Economic Review*, *104*(1), 291–307.

Kremer, M. (1993). The o-ring theory of economic development. *Quarterly Journal of Economics*, *108*(3), 551–575.

Lazear, E. P. y Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, *21*(4), 91–114.

Legros, P. y Newman, A. F. (2007). Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica*, *75*(4), 1073–1102.

Marschak, J. y Radner, R. (1972). *Economic Theory of Teams*. New Haven, Conn.: Yale University Press.

McAfee, R. P. y McMillan, J. (1991). Optimal contracts for teams. *International Economic Review*, *32*(3), 561–577.

Mookherjee, D. (2006). Decentralization, hierarchies, and incentives: A mechanism design perspective. Journal of Economic Literature, 44(2), 367–390.

Poitevin, M. (2000). Can the theory of incentives explain decentralization? Canadian Journal of Economics / Revue canadienne d'Economique, 33(4), 878–906.

Prat, A. (2002). Should a team be homogeneous? European Economic Review, 46(7), 1187–1207.

Sappington, D. (1983). Limited liability contracts between principal and agent. Journal of Economic Theory, 29(1), 1 – 21.

Sappington, D. (1984). Incentive contracting with asymmetric and imperfect precontractual knowledge. Journal of Economic Theory, 34(1), 52 – 70.

Serfes, K. (2005). Risk sharing vs. incentives: Contract design under two-sided heterogeneity. Economics Letters, 88(3), 343 – 349.

Serfes, K. (2007). Endogenous matching in a market with heterogeneous principals and agents. International Journal of Game Theory, 36(3-4), 587–619.

Shavell, S. (1979). On moral hazard and insurance. Quarterly Journal of Economics, 93(4), 541–562.

Shimer, R. y Smith, L. (2000). Assortative matching and search. Econometrica, 68(2), 343–369.

Smith, L. (2006). The marriage model with search frictions. Journal of Political Economy, 114(6), 1124–1144.

Winter, E. (2004). Incentives and discrimination. American Economic Review, 94(3), 764–773.

Wright, D. J. (2004). The risk and incentives trade-off in the presence of heterogeneous managers. Journal of Economics, 83(3), 209–223.

# A Proofs

## A.1 Proof of Lemma 1

Take an arbitrary incentive feasible contract $(\mu, w)$. Form a partition of workers $\Pi := \{\pi_1, ..., \pi_K\}$, with $|\pi_i| = N_i > 0$, in which workers in the same partition element $\pi_i$ face the same interim probabilities, $(p_H^{\pi_i}, p_L^{\pi_i})$, under $\mu$. Now, define a set of workers $I := \{\iota_1, \iota_2, ..., \iota_K\}$ such that $\iota_k \in \pi_k$ and expected wage payments to $\iota_k$ are minimized across all workers in $\pi_k$.

To construct an $\mathrm{independent}$ wage scheme, for each partition element $\pi_k$ and every worker $i \in \pi_k$, if $i$'s type is $t$, his teammate's type is $t'$, and his team produces output $y$, then set his wage equal to the expected wage given to $\iota_k$ when $\iota_k$'s type is $t$, his teammate's type is $t'$, and his team produces output $y$. By construction, this scheme satisfies all incentive constraints given $\mu$ and weakly decreases the manager's expected wage payments. Abusing notation, let it be denoted by $(w_t^{t'}(y, \iota))_{t,t' \in T, y \in Y, \iota \in I}$.

Now, before applying the matching $\mu$, uniformly randomize over the set of all permutations of the $N$ workers. In the resulting matching $\hat{\mu}$, each worker has interim probabilities $p_H^{\hat{\mu}} = \sum_{i=1}^K \frac{N_i}{N} p_H^{\pi_i}$ and $p_L^{\hat{\mu}} = \sum_{i=1}^K \frac{N_i}{N} p_L^{\pi_i}$ so that $\hat{\mu}$ satisfies equal treatment. Take the corresponding expectation over wages, i.e. if a worker's type is $t$, his teammate's type is $t'$ and his team produces output $y$, then set his wage equal to,

$$\hat{w}_t^{t'}(y) = \sum_{\iota \in I} \frac{N_i}{N} w_t^{t'}(y, \iota).$$

Call the resulting wage scheme $\hat{w}$. By construction, $(\hat{\mu}, \hat{w})$ is incentive feasible, expected output under $\hat{\mu}$ is identical to expected output under $\mu$, and expected wage payments under $w$ are at least as large as under $\hat{w}$. Hence, $(\hat{\mu}, \hat{w})$ attains at least the same profits as $(\mu, w)$.[27]

[27] We thank Juuso Toikka for suggesting this short proof. A longer proof establishing that equal treatment matchings strictly outperform all non-equal treatment matchings when RM is optimal is available upon request.

## A.2 Proof of Lemma 2

We first show that $w_H^H(0) = w_H^L(0) = 0$ in any optimal anonymous, independent wage scheme $w := (w_t^{t'}(y))$. There are two cases to consider. First, consider the case in which $\bar{u}_L = 0$ at the optimal wage scheme. In this case, if either $w_H^H(0) > 0$ or $w_H^L(0) > 0$, $IC_L$ would be violated; lows could misreport their type, never exert effort, and attain strictly positive expected utility. Second, consider the case in which $\bar{u}_L > 0$ at the optimal wage scheme. In this case, at least one of the following conditions must be satisfied: (i) $w_L^H(1) - w_L^H(0) > \frac{c}{q_M}$, (ii) $w_L^L(1) - w_L^L(0) > \frac{c}{q_L}$, (iii) $w_L^L(0) > 0$, or (iv) $w_L^H(0) > 0$. We claim that if either $w_H^H(0) > 0$ or $w_H^L(0) > 0$, then we can construct an alternative wage scheme $\hat{w} := (\hat{w}_t^{t'}(y))$ which yields the manager strictly higher profits, contradicting the supposed optimality of $w$. Suppose $w_H^H(0) > 0$. Then, we can construct $\hat{w}$ as follows. Modify $w$ so that $\hat{w}_H^H(0) = 0$ and $\hat{w}_H^H(1) = w_H^H(1) + \frac{(1-q_H)}{q_H} w_H^H(0)$. Then, $\bar{u}_H$ is unchanged, all incentive constraints for highs are satisfied, and the value of the right-hand side of $IC_L$ strictly decreases. Hence, there exists an $\epsilon > 0$ by which we may reduce some wage payment to lows and satisfy all obedience and incentive constraints: in case $(i)$, set $\hat{w}_L^H(1) = w_L^H(1) - \epsilon$; in case $(ii)$, set $\hat{w}_L^L(1) = w_L^L(1) - \epsilon$; in case $(iii)$, set $\hat{w}_L^L(0) = w_L^L(0) - \epsilon$; and in case $(iv)$, set $\hat{w}_L^H(0) = w_L^H(0) - \epsilon$. It follows that $\hat{w}$ strictly increases the manager's profits and so it must be the case that $w_H^H(0) = 0$ in any optimal wage scheme. A similar argument shows that $w_H^L(0) > 0$ as well.

To show that $IC_H$ must bind in any optimal wage scheme, we first show that either $O_H^L$ or $O_H^H$ must be slack. Since $w_H^H(0) = w_H^L(0) = 0$ in any optimal wage scheme, if both $O_H^L$ and $O_H^H$ bind, then $\bar{u}_H = 0$. On the other hand, the payoff to deviating is,

$$\begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_L^H(1) + (1 - q_H) w_L^H(0) - c \\ q_M w_L^L(1) + (1 - q_M) w_L^L(0) - c \end{bmatrix} \geq \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} (\frac{q_H}{q_M} - 1)c + w_L^H(0) \\ (\frac{q_M}{q_L} - 1)c + w_L^L(0) \end{bmatrix},$$

where the inequality follows because $O_L^H$ and $O_L^L$ must be satisfied in any incentive feasible wage scheme. But since $\frac{q_H}{q_M} - 1 > 0$ and $\frac{q_M}{q_L} - 1 > 0$, the right-hand side of the inequality must be strictly positive. As optimal wage schemes must be feasible, it follows that either $O_H^H$ or $O_H^L$ must be slack.

Now, towards contradiction, suppose that $IC_H$ is slack at the optimal wage scheme $w$. Since either $O_H^H$ or $O_H^L$ must be slack, we may then reduce $w_H^H(1)$ or $w_L^H(1)$ by a small amount (depending on whether $O_H^H$ or $O_H^L$ is slack) and still satisfy all incentive constraints for highs. This modification does not affect $O_L^H$ and $O_L^L$ and weakly decreases the right-hand side of $IC_L$. Hence, all incentive constraints for lows remain satisfied as well. As the manager's profits must strictly increase in this modified incentive feasible contract, $w$ could not have been optimal, our desired contradiction.

Finally, we show that $O_L^H$ and $O_L^L$ must bind. Towards contradiction, suppose that $O_L^H$ does not bind at an optimal contract $w$. Consider a modified contract $\hat{w}$, where, for a small $\epsilon > 0$, $\hat{w}_L^H(0) = w_L^H(0) + \epsilon$. Then, we may set $\hat{w}_L^H(1) = w_L^H(1) - \frac{(1-q_M)}{q_M}\epsilon$, so that $\bar{u}_L$ is unchanged. Notice that if $\epsilon$ is small enough, however, $O_L^H$ remains slack. Further, the right-hand side of $IC_H$ strictly decreases:

$$\begin{aligned} q_H \hat{w}_L^H(1) + (1-q_H)\hat{w}_L^H(0) &= q_H w_L^H(1) + (1-q_H)\hat{w}_L^H(0) - \frac{q_H - q_M}{q_M}\epsilon \\ &< q_H w_L^H(1) + (1-q_H) w_L^H(0). \end{aligned}$$

Since either $O_H^H$ or $O_H^L$ is slack in any optimal contract, as shown previously, we may then reduce $w_H^H(1)$ or $w_H^L(1)$ by a small amount and strictly increase the manager's profits while satisfying all incentive constraints, contradicting the supposed optimality of $w$. A similar argument shows that $O_L^L$ binds as well.

### A.3 Proof of Theorem 1

#### Simplifying the Manager's Minimization Problem

We first simplify the manager's objective function. Let $p^{\mu}_{(H,H)} = pp^{\mu}_{H}\frac{N}{2}$, $p^{\mu}_{(L,L)} = (1-p)p^{\mu}_{L}\frac{N}{2}$, and $p^{\mu}_{(H,L)} = ((1-p)p^{\mu}_{L}+p(1-p^{\mu}_{H}))\frac{N}{2}$ denote the expected number of teams composed of two highs, a high and a low, and two lows. In any matching, it must be the case that the ex-ante probability a worker is a low and matches a high equals the ex-ante probability a worker is a high and matches a low, i.e. $(1-p)p^{\mu}_{L} = p(1-p^{\mu}_{H})$. Hence, $p^{\mu}_{(H,L)} = ((1-p)p^{\mu}_{L}+p(1-p^{\mu}_{H}))\frac{N}{2} = p(1-p^{\mu}_{H})N = (1-p)p^{\mu}_{L}N$. The manager's expected wage payments are therefore given by,

$$C(\mu) = \underbrace{p^{\mu}_{(H,H)}\Big[q_H(2w^H_H(1)) + (1-q_H)(2w^H_H(0))\Big] + p^{\mu}_{(H,L)}\Big[q_M w^L_H(1) + (1-q_M)w^L_H(0)\Big]}_{\text{Expected Wage Payments to Highs}} +$$

$$\underbrace{p^{\mu}_{(L,L)}\Big[q_L(2w^L_L(1)) + (1-q_L)(2w^L_L(0))\Big] + p^{\mu}_{(H,L)}\Big[q_M w^H_L(1) + (1-q_M)w^H_L(0)\Big]}_{\text{Expected Wage Payments to Lows}}$$

$$= N(p\bar{u}_H + (1-p)\bar{u}_L + c).$$

As $N$ and $c$ are positive, it suffices to consider choosing wages to minimize the expected information rent paid per worker,

$$\gamma(\mu) := p\bar{u}_H + (1-p)\bar{u}_L.$$

We now utilize Lemma 2 to re-write the constraints of the manager's problem in terms of wage payments to highs, $w^H_H(1)$ and $w^L_H(1)$, and the utility payment to lows $\bar{u}_L$. By Lemma 2, $w^H_H(0) = w^L_H(0) = 0$ in any optimal wage scheme. Hence, $IC_H$ simplifies to

$$\begin{bmatrix} p^{\mu}_{H} \\ 1-p^{\mu}_{H} \end{bmatrix}^T \begin{bmatrix} q_H w^H_H(1) - c \\ q_M w^L_H(1) - c \end{bmatrix} \geq \begin{bmatrix} p^{\mu}_{L} \\ 1-p^{\mu}_{L} \end{bmatrix}^T \begin{bmatrix} q_H w^H_L(1) + (1-q_H)w^H_L(0) - c \\ q_M w^L_L(1) + (1-q_M)w^L_L(0) - c \end{bmatrix}.$$

Further, by Lemma 2, we know that obedience constraints $O^H_L$ and $O^L_L$ bind in any optimal wage scheme. These constraints may be written as difference equations $w_L^H(1) - w_L^H(0) = \frac{c}{q_M}$ and $w_L^L(1) - w_L^L(0) = \frac{c}{q_L}$. Substituting them into $IC_H$, we obtain

$$\begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} \geq \underbrace{\begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} w_L^H(0) \\ w_L^L(0) \end{bmatrix}}_{=\bar{u}_L} + \begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix}.$$

The simplified manager's problem, written in terms of $\bar{u}_L$, $w_H^H(1)$, and $w_H^L(1)$ is therefore,

$$\min_{(\bar{u}_L, w_H^H(1), w_H^L(1)) \in \mathbb{R}_+^3} p \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} + (1-p)\bar{u}_L$$

subject to

$$[IC_H] \quad \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} \geq \bar{u}_L + \begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix}$$

$$[IC_L] \quad \bar{u}_L \geq \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} \max\{q_M w_H^H(1) - c, 0\} \\ \max\{q_L w_H^L(1) - c, 0)\} \end{bmatrix}$$

$$[OB_H^H] \quad q_H w_H^H(1) - c \geq 0$$

$$[OB_H^L] \quad q_M w_H^L(1) - c \geq 0.$$

**Case 1: $\mu$ exhibits positive (negative) assortativity and the production technology is log supermodular (submodular).**

The derivation in Section 4.2.1 establishes that, when $w_L^L(0) = w_L^H(0) = 0$, $w_L^L(1) = \frac{c}{q_L}$ and $w_L^H(1) = \frac{c}{q_M}$, so that $\bar{u}_L = 0$, there exist wages $w_H^H(1)$ and $w_H^L(1)$ at which both $IC_H$ and $IC_L$ are satisfied. By Lemma 2, $IC_H$ must bind at any optimal wage scheme. Observing that the slope of the manager's isocost curve and that of $IC_H$ coincide, the manager is indifferent between any wages on the

$IC_H$ line. The optimal wages for highs are therefore $w_H^H(0) = w_H^L(0) = 0$ and any $w_H^H(1) \geq 0$ and $w_H^L(1) \geq 0$ satisfying,

$$\bar{u}_H = \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} = \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c - c \\ \frac{q_M}{q_L} c - c \end{bmatrix} > 0.$$

**Case 2: $\mu$ exhibits positive (negative) assortativity and the production technology is strictly log submodular (supermodular).**

The derivation in Section 4.2.1 establishes that, in this case, there is no wage scheme satisfying $IC_H$ and $IC_L$ that has $\bar{u}_L = 0$. By the binding constraints $O_L^H$ and $O_L^L$, this immediately implies that either $w_L^L(0) > 0$ or $w_L^H(0) > 0$.

To find the optimal wages, we simplify the manager's problem further by eliminating $\bar{u}_L$, so that she need only choose $w_H^H(1)$ and $w_H^L(1)$. To do this, observe that, since $IC_H$ binds in any optimal wage scheme,

$$\bar{u}_L = \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} - \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c - c \\ \frac{q_M}{q_L} c - c \end{bmatrix}. \qquad (IC_H')$$

Substituting this expression into $IC_L$, and into the objective function, we obtain the minimization problem,

$$\min_{(w_H^H(1), w_H^L(1)) \in \mathbb{R}_+^2} \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} - (1-p) \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c - c \\ \frac{q_M}{q_L} c - c \end{bmatrix}$$

subject to

$$[IC_L] \quad \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} - \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M} c - c \\ \frac{q_M}{q_L} c - c \end{bmatrix} \geq \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} \max\{q_M w_H^H(1) - c, 0\} \\ \max\{q_L w_H^L(1) - c, 0\} \end{bmatrix}$$

$$[OB_H^H] \quad q_H w_H^H(1) - c \geq 0$$

$$[OB_H^L] \quad q_M w_H^L(1) - c \geq 0.$$

We now show that we can eliminate the "max" operators on the right-hand side of $IC_L$.

**Claim 1.** *In any optimal wage scheme,* $w_H^H(1) \geq \frac{c}{q_M}$ *and* $w_H^L(1) \geq \frac{c}{q_L}$.

*Proof.* Suppose, towards contradiction, that $w_H^L(1) < \frac{c}{q_L}$ in some optimal wage scheme. Then, to satisfy $IC_L$, it must be the case that $w_H^H(1) > \frac{c}{q_M}$. $IC_L$ thus simplifies to,

$$\begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} (q_H - q_M)w_H^H(1) \\ q_M w_H^L(1) - c \end{bmatrix} \geq \begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix}.$$

Now, decrease $w_H^H(1)$ and increase $w_H^L(1)$ until the left-hand side attains its original value. Inspecting the manager's objective function, this strictly decreases expected wage payments, contradicting the supposed optimality of the original wage scheme. A similar proof shows that $w_H^H(1) \geq \frac{c}{q_M}$. □

Eliminating the constant from the manager's objective function and the max operators from $IC_L$, we obtain the following problem:

$$\min_{(w_H^H(1), w_H^L(1)) \in \mathbb{R}_+^2} \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix}$$

subject to

$$[IC_L] \quad \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} (q_H - q_M)w_H^H(1) \\ (q_M - q_L)w_H^L(1) \end{bmatrix} \geq \begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix}$$

$$[OB_H^H] \quad q_H w_H^H(1) - c \geq 0$$

$$[OB_H^L] \quad q_M w_H^L(1) - c \geq 0.$$

Notice, the left-hand side of $IC_L$ is increasing in $w_H^H(1)$ and $w_H^L(1)$ and the constraint is not satisfied when $OB_H^H$ and $OB_H^L$ bind. Hence, to minimize wage payments, the manager chooses $w_H^H(1)$ and $w_H^L(1)$ so that the constraint holds with equality, i.e. $IC_L$ binds.

We now identify the optimal values of $w_H^H(1)$ and $w_H^L(1)$ to pin down all optimal wages. Notice, the slope of the manager's isocost line when written with $w_H^H(1)$ on the left-hand side is,

$$-\frac{1-p_H^\mu}{p_H^\mu}\frac{q_M}{q_H},$$

while the slope of the worker's incentive constraint (which holds with equality at the optimal wage scheme) when written with $w_H^H(1)$ on the left-hand side is,

$$-\frac{1-p_H^\mu}{p_H^\mu}\frac{q_M-q_L}{q_H-q_M}.$$

When the slope of the isocost line is more negative than that of the worker's incentive constraint, then setting $w_H^L(1)=\frac{c}{q_L}$ and using $IC_L$ to determine $w_H^H(1)$ is optimal. This happens if and only if,

$$-\frac{1-p_H^\mu}{p_H^\mu}\frac{q_M}{q_H}\leq -\frac{1-p_H^\mu}{p_H^\mu}\frac{q_M-q_L}{q_H-q_M} \iff \frac{q_M}{q_H}\geq\frac{q_M-q_L}{q_H-q_M} \iff \frac{q_H}{q_M}\geq\frac{q_M}{q_L}.$$

Put differently, if $\mu$ exhibits negative assortativity and the technology is log supermodular, then $w_H^L(1)=\frac{c}{q_L}$ is optimal. Using the binding constraint $IC_L$, the optimal value of $w_H^H(1)$ is given by,

$$w_H^H(1)=-\frac{1-p_H^\mu}{p_H^\mu}\frac{q_M-q_L}{q_H-q_M}\frac{c}{q_L}+\frac{1}{p_H^\mu(q_H-q_M)}\begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T\begin{bmatrix} \frac{q_H}{q_M}c-c \\ \frac{q_M}{q_L}c-c \end{bmatrix}.$$

Similarly, if $\mu$ exhibits positive assortativity and the technology is log submodular, then $w_H^H(1)=\frac{c}{q_M}$ is optimal. Using the binding constraint $IC_L$, the optimal value of $w_H^L(1)$ is given by,

$$w_H^L(1)=-\frac{p_H^\mu}{1-p_H^\mu}\frac{q_H-q_M}{q_M-q_L}\frac{c}{q_M}+\frac{1}{(1-p_H^\mu)(q_M-q_L)}\begin{bmatrix} p_L^\mu \\ 1-p_L^\mu \end{bmatrix}^T\begin{bmatrix} \frac{q_H}{q_M}c-c \\ \frac{q_M}{q_L}c-c \end{bmatrix}.$$

Once the manager has found the optimal wages $(w_H^H(1), w_H^L(1))$, she pins down the optimal value of $\bar{u}_L^*$ at any optimal contract using the binding constraint $IC'_H$. As the manager's expected payments are proportional to $\bar{u}_L$, any wages for lows such that $O_L^H$ and $O_L^L$ bind and $\bar{u}_L = \bar{u}_L^*$ are optimal.

## A.4 Proof of Theorem 2

Total expected output is given by,

$$\begin{aligned} O(\mu) =& p^{\mu}_{(H,H)} q_H + p^{\mu}_{(H,L)} q_M + p^{\mu}_{(L,L)} q_L \\ =& \frac{N}{2}\left[ p p_H^{\mu} q_H + (p(1-p_H^{\mu}) + (1-p)p_L^{\mu}) q_M + (1-p)(1-p_L^{\mu}) q_L \right], \end{aligned}$$

where $p^{\mu}_{(H,H)} = p p_H^{\mu} \frac{N}{2}$, $p^{\mu}_{(L,L)} = (1-p)p_L^{\mu} \frac{N}{2}$, and $p^{\mu}_{(H,L)} = ((1-p)p_L^{\mu} + p(1-p_H^{\mu}))\frac{N}{2}$ denote the expected number of teams composed of two highs, a high and a low, and two lows. From the proof of Theorem 1, we saw that the manager's total expected wage payments are given by,

$$C(\mu) = N(p\bar{u}_H + (1-p)\bar{u}_L + c).$$

Let $C^*(\mu)$ denote its value at an optimal wage scheme given $\mu$, and $\bar{u}_H^*$ and $\bar{u}_L^*$ the corresponding information rents paid to each worker. Then, the manager's profit maximizing problem is to choose an equal treatment matching $\mu$ to maximize,

$$\begin{aligned} O(\mu) - C^*(\mu) =& \frac{N}{2}\left[ p p_H^{\mu} q_H + (p(1-p_H^{\mu}) + (1-p)p_L^{\mu}) q_M + (1-p)(1-p_L^{\mu}) q_L \right] \\ & - N(p\bar{u}_H^* + (1-p)\bar{u}_L^* + c). \end{aligned}$$

As $N$ is positive and $Nc$ does not depend on $\mu$, it is therefore equivalent to choose an equal treatment matching $\mu$ to maximize

$$o(\mu) - \gamma^*(\mu),$$

where $o(\mu) := O(\mu)/N$ is the expected output produced per worker and $\gamma^*(\mu) := p\bar{u}_H^* + (1-p)\bar{u}_L^*$ is the optimal expected information rent paid per worker. We now show that both $o(\mu)$ and $\gamma^*(\mu)$ depend only on $\mu$ through $p_H^\mu$ and provide comparative statics for each in terms of this parameter.

**Claim 2.** *$o(\mu)$ depends on $\mu$ only through $p_H^\mu$, and is linear and strictly increasing in $p_H^\mu$.*

*Proof.* Observe that,

$$
\begin{aligned}
o(\mu) =& \frac{1}{2}\left[p p_H^\mu q_H + (p(1-p_H^\mu) + (1-p)p_L^\mu) q_M + (1-p)(1-p_L^\mu) q_L\right] \\
=& p_H^\mu \frac{p}{2}(q_H + q_L - 2q_M) + p q_M + \frac{1-2p}{2} q_L,
\end{aligned} \tag{1}
$$

where the inequality follows by imposing the requirement that $p(1-p_H^\mu) = (1-p)p_L^\mu$ so that $p_L^\mu = \frac{p}{1-p}(1-p_H^\mu)$. Clearly, the expression is linear in $p_H^\mu$. As expected output satisfies strictly increasing differences, $q_H + q_L - 2q_M > 0$. Hence, $o(\mu)$ is strictly increasing in $p_H^\mu$. □

**Claim 3.** *$\gamma^*(\mu)$ depends on $\mu$ only through $p_H^\mu$, and is piecewise linear, continuous and convex in $p_H^\mu$. If the production technology is strictly log supermodular (submodular), then $\gamma^*(\mu)$ is strictly decreasing (increasing) in $p_H^\mu$. If the production technology is strictly log supermodular, then the slope of $\gamma^*(\mu)$ is equal to*

$$
\frac{\partial \gamma^*(\mu)}{\partial p_H^\mu} = \begin{cases} c\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{c}{1-p}\frac{q_M^2 - q_H q_L}{q_M(q_M - q_L)} & \text{if } p_H^\mu < p \\ c\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) & \text{if } p_H^\mu > p, \end{cases}
$$

*and if the production technology is strictly log submodular, then the slope of $\gamma^*(\mu)$ is equal to*

$$
\frac{\partial \gamma^*(\mu)}{\partial p_H^\mu} = \begin{cases} c\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) & \text{if } p_H^\mu < p \\ c\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{c}{1-p}\frac{q_M^2 - q_H q_L}{q_L(q_H - q_M)} & \text{if } p_H^\mu > p, \end{cases}
$$

*Finally, if the production technology is both log submodular and log supermodular, then $\gamma^*(\mu)$ is constant in $p_H^\mu$.*

*Proof.* If $p_H^\mu > p$ ($p_H^\mu < p$) and the production technology is log supermodular (submodular), then $\bar{u}_L^* = 0$ and

$$\bar{u}_H^* = \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix} = p_H^\mu \frac{cp}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \alpha_1, \tag{2}$$

where $\alpha_1 := \frac{p}{1-p}\frac{q_H}{q_M}c + \frac{1-2p}{1-p}\frac{q_M}{q_L}c - c$ is a constant that does not depend on $\mu$ and $p_L^\mu = \frac{p}{1-p}(1 - p_H^\mu)$ is applied to obtain the second equality. Consequently,

$$\gamma^*(\mu) = p\bar{u}_H^* + (1-p)\bar{u}_L^* = p_H^\mu \frac{cp^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + p\alpha_1, \tag{3}$$

which depends on $\mu$ only through $p_H^\mu$. If the production technology is both log submodular and log supermodular, then $\frac{q_H}{q_M} = \frac{q_M}{q_L}$ so that $\gamma^*(\mu)$ is constant in $p_H^\mu$. Under strict log supermodularity (submodularity), $\gamma^*(\mu)$ is linear and strictly decreasing (increasing) in $p_H^\mu$ on $(p, 1)$ ($(0, p)$), with

$$\frac{\partial \gamma^*(\mu)}{\partial p_H^\mu} = c\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right).$$

For the other cases, as seen in the proof of Theorem 1,

$$\bar{u}_H^* = \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix},$$

and,

$$\bar{u}_L^* = \begin{bmatrix} p_H^\mu \\ 1 - p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) - c \\ q_M w_H^L(1) - c \end{bmatrix} - \begin{bmatrix} p_L^\mu \\ 1 - p_L^\mu \end{bmatrix}^T \begin{bmatrix} \frac{q_H}{q_M}c - c \\ \frac{q_M}{q_L}c - c \end{bmatrix},$$

where $w_H^H(1)$ and $w_H^L(1)$ are optimal wages. Hence,

$$\begin{aligned}\gamma^*(\mu) &= p\bar{u}_H^* + (1-p)\bar{u}_L^* \\ &= \begin{bmatrix} p_H^\mu \\ 1-p_H^\mu \end{bmatrix}^T \begin{bmatrix} q_H w_H^H(1) \\ q_M w_H^L(1) \end{bmatrix} + p_H^\mu c p \left(\frac{q_H}{q_M} - \frac{q_M}{q_L}\right) + \alpha_2,\end{aligned}$$

where the second equality follows from the second equality in Equation 2 and

$$\alpha_2 := -(1-p)\alpha_1 - c = p\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right)c + p\left(\frac{q_M}{q_L} - 1\right)c - \frac{q_M}{q_L}c$$

is a constant that does not depend on $\mu$.

If the production technology is strictly log supermodular and $p_H^\mu < p$, then $w_H^H(1) = \frac{c}{q_M}$, so that

$$p_H^\mu q_H w_H^H(1) = p_H^\mu c \frac{q_H}{q_M}.$$

Further,

$$w_H^L(1) = -\frac{p_H^\mu}{1-p_H^\mu}\frac{q_H - q_M}{q_M - q_L}\frac{c}{q_M} + \frac{1}{(1-p_H^\mu)(q_M - q_L)}\left(p_H^\mu \frac{cp}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \alpha_1\right),$$

so that,

$$(1-p_H^\mu) q_M w_H^L(1) = p_H^\mu c\left[\frac{p}{1-p}\frac{q_M^2 - q_H q_L}{q_L(q_M - q_L)} - \frac{q_H - q_M}{q_M - q_L}\right] + \alpha_1 \frac{q_M}{q_M - q_L}.$$

Consequently,

$$\gamma^*(\mu) = p_H^\mu c\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{1}{1-p}\frac{q_M^2 - q_H q_L}{q_M(q_M - q_L)}\right] + \alpha_3,$$

where

$$\alpha_3 := \alpha_2 + \alpha_1 \frac{q_M}{q_M - q_L}$$

is a constant that does not depend on $\mu$. Once again, the resulting expression depends on $\mu$ only through $p_H^\mu$. Further, if $p_H^\mu = p$, we obtain the expression

$$\gamma^*(\mu) = cp\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right)\right] + p\alpha_1,$$

so that $\gamma^*(\mu)$ is continuous on $[0,1]$ when the production technology satisfies strict log supermodularity. Finally, $\gamma^*(\mu)$ is linear and strictly decreasing in $(0,p)$, with slope equal to

$$\frac{\partial \gamma^*(\mu)}{\partial p_H^\mu} = c\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{1}{1-p}\frac{q_M^2 - q_H q_L}{q_M(q_M - q_L)}\right] < 0,$$

where the inequality follows because $\frac{q_H}{q_M} > \frac{q_M}{q_L}$ implies $q_M^2 < q_H q_L$.

We now consider the case in which the production technology is strictly log submodular and $p_H^\mu > p$. Then, $w_H^L(1) = \frac{c}{q_L}$, so that

$$(1 - p_H^\mu) q_M w_H^L(1) = (1 - p_H^\mu) c \frac{q_M}{q_L}.$$

Hence,

$$\gamma^*(\mu) = p_H^\mu w_H^H(1) + p_H^\mu c\left[p\left(\frac{q_H}{q_M} - \frac{q_M}{q_L}\right) - \frac{q_M}{q_L}\right] + (1 - p_H^\mu) q_M w_H^L(1) + \alpha_4,$$

with

$$\alpha_4 := \alpha_2 + c\frac{q_M}{q_L} = p\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right)c + p\left(\frac{q_M}{q_L} - 1\right)c.$$

Further,

$$w_H^H(1) = -\frac{1 - p_H^\mu}{p_H^\mu}\frac{q_M - q_L}{q_H - q_M}\frac{c}{q_L} + \frac{1}{p_H^\mu(q_H - q_M)}\left(p_H^\mu \frac{cp}{1-p}\left(\frac{q_H}{q_M} - \frac{q_M}{q_L}\right) + \alpha_1\right),$$

so that,

$$p_H^\mu q_H w_H^H(1) = p_H^\mu c\left[\frac{p}{1-p}\frac{q_H(q_H q_L - q_M^2)}{q_L q_M(q_H - q_M)} + \frac{q_H(q_M - q_L)}{q_L(q_H - q_M)}\right] + \alpha_1 \frac{q_H}{q_H - q_M} - c\frac{q_H(q_M - q_L)}{q_L(q_H - q_M)}.$$

Consequently,

$$\gamma^*(\mu) = p_H^\mu c\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{1}{1-p}\frac{q_M^2 - q_H q_L}{q_L(q_H - q_M)}\right] + \alpha_5 \tag{4}$$

where

$$\alpha_5 := \alpha_4 + \alpha_1 \frac{q_H}{q_H - q_M} - c\frac{q_H(q_M - q_L)}{q_L(q_H - q_M)},$$

is a constant that does not depend on $\mu$. Further, if $p_H^\mu = p$, then

$$\gamma^*(\mu) = cp\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right)\right] + p\alpha_1,$$

so that $\gamma^*(\mu)$ is continues when the production function is strictly log submodular. Finally, $\gamma^*(\mu)$ is linear and strictly increasing in $(p, 1)$, with slope equal to

$$\frac{\partial \gamma^*(\mu)}{\partial p_H^\mu} = c\left[\frac{p^2}{1-p}\left(\frac{q_M}{q_L} - \frac{q_H}{q_M}\right) + \frac{1}{1-p}\frac{q_M^2 - q_H q_L}{q_L(q_H - q_M)}\right] > 0,$$

where the inequality follows because $\frac{q_M}{q_L} > \frac{q_H}{q_M}$ implies $q_M^2 - q_H q_L > 0$. □

We put together the two claims to prove the Theorem. If the production technology is log supermodular, then $\gamma^*(\mu)$ is decreasing, at least weakly, in $p_H^\mu$. As $o(\mu)$ is strictly increasing in $p_H^\mu$, it is optimal to maximize $p_H^\mu$, i.e. PAM is optimal.

If the production technology is strictly log submodular, however, Claim 3 implies that $\gamma^*(\mu)$ is strictly increasing in $p_H^\mu$ and piecewise linear convex with a kink at $p$. And since $o(\mu)$ is linear, $o(\mu) - \gamma^*(\mu)$ is piecewise linear concave with a kink at $p$. Therefore, PAM is the unique optimal matching if and only if $o(\mu) - \gamma(\mu)$ is strictly increasing in $p_H^\mu$ on $(p, 1)$, NAM is the unique optimal matching if and only if $o(\mu) - \gamma(\mu)$ is strictly decreasing in $p_H^\mu$ on $(0, p)$, and RM is the unique optimal matching if and only if $o(\mu) - \gamma(\mu)$ is strictly increasing in $p_H^\mu$ on $(0, p)$ and strictly decreasing in $p_H^\mu$ on $(p, 1)$. As profits are differentiable on $(p,1)$, profits are strictly increasing in $p_H^\mu$ on $(p,1)$ if and only if,

$$\frac{\partial}{\partial p_H^\mu}[o(\mu)-\gamma(\mu)] = \frac{p}{2}(q_H+q_L-2q_M)-c\left[\frac{1}{1-p}\frac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}+\frac{p^2}{1-p}\left(\frac{q_M}{q_L}-\frac{q_H}{q_M}\right)\right]>0,$$

or, when,

$$0<c<\underline{c}:=\frac{1}{2}\left(\frac{1-p}{p}\right)\left(\frac{q_H+q_L-2q_M}{(\frac{q_M}{q_L}-\frac{q_H}{q_M})+\frac{1}{p^2}\left(\frac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}\right)}\right).$$

As profits are differentiable on $(0,p)$, profits are strictly decreasing in $p_H^\mu$ on $(0,p)$ if and only if,

$$\frac{\partial}{\partial p_H^\mu}[o(\mu)-\gamma(\mu)] = \frac{p}{2}(q_H+q_L-2q_M)-c\frac{p^2}{1-p}\left(\frac{q_M}{q_L}-\frac{q_H}{q_M}\right)<0,$$

or, when,

$$c>\bar{c}:=\frac{1}{2}\left(\frac{1-p}{p}\right)\left(\frac{q_H+q_L-2q_M}{\frac{q_M}{q_L}-\frac{q_H}{q_M}}\right)>0.$$

Hence, PAM is the unique optimal matching if $c<\underline{c}$, NAM is the unique optimal matching if $c>\bar{c}$, and RM is the unique optimal matching if $\underline{c}<c<\bar{c}$.

### A.5 Proof of Theorem 3

Define $p_H^{PAM} := \max_{\mu\in\mathcal{M}} \; p_H^\mu$ and $p_H^{NAM} := \min_{\mu\in\mathcal{M}} \; p_H^\mu$, where $\mathcal{M}$ is the set of all equal treatment matchings. By Lemma 4, it is optimal to pay each worker $\frac{c}{q_L}$ if their team produces high output and $0$ if their team produces low output and implement PAM. For analytical convenience, we write the expected information rent paid per worker under delegation in terms of $p_L^{PAM}$ instead of $p_H^{PAM}$, using the observation that $1-p_H^{PAM}=\frac{1-p}{p}p_L^{PAM}$,

$$\begin{aligned}\gamma^D :=& cp\left[p_H^{PAM}\frac{q_H}{q_L}+(1-p_H^{PAM})\frac{q_M}{q_L}-1\right]+c(1-p)\left[p_L^{PAM}\frac{q_M}{q_L}+(1-p_L^{PAM})-1\right]\\ =& cp\left[\frac{q_H}{q_L}+\frac{1-p}{p}p_L^{PAM}\frac{2q_M-q_H-q_L}{q_L}-1\right],\end{aligned}$$

Using Equation 4 in the proof of Theorem 1, we also write the expected information rent paid per worker under PAM in terms of $p_L^{PAM}$ instead of $p_H^{PAM}$,

$$\begin{aligned}
\gamma^{PAM} :=& \frac{cp}{1-p}\left[p_H^{PAM}\left(p\left(\frac{q_M}{q_L}-\frac{q_H}{q_M}\right)+\frac{1}{p}\frac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}\right)+p\left(\frac{q_H}{q_M}-\frac{q_M}{q_L}\right)+\frac{q_Hq_L-q_M^2}{q_L(q_H-q_M)}+(1-p)\left(\frac{q_M}{q_L}-1\right)\right]\\
=& \frac{cp}{1-p}\left[\frac{1-p}{p}p_L^{PAM}\left(p\left(\frac{q_H}{q_M}-\frac{q_M}{q_L}\right)+\frac{1}{p}\frac{q_Hq_L-q_M^2}{q_L(q_H-q_M)}\right)+\frac{1-p}{p}\frac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}+(1-p)\left(\frac{q_M}{q_L}-1\right)\right]\\
=& c\left[p_L^{PAM}\left(p\left(\frac{q_H}{q_M}-\frac{q_M}{q_L}\right)+\frac{1}{p}\frac{q_Hq_L-q_M^2}{q_L(q_H-q_M)}\right)+\frac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}+p\left(\frac{q_M}{q_L}-1\right)\right].
\end{aligned}$$

Expected wage payments under PAM are larger than under delegation if and only if $\gamma^D < \gamma^{PAM}$, i.e.

$$\begin{aligned}
& c\left[p\tfrac{q_H}{q_L}+(1-p)p_L^{PAM}\tfrac{2q_M-q_H-q_L}{q_L}-p\right] < c\left[p_L^{PAM}\left(p\left(\tfrac{q_H}{q_M}-\tfrac{q_M}{q_L}\right)+\tfrac{1}{p}\tfrac{q_Hq_L-q_M^2}{q_L(q_H-q_M)}\right)+\tfrac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}+p\left(\tfrac{q_M}{q_L}-1\right)\right]\\
\iff & p_L^{PAM}\left((1-p)\tfrac{2q_M-q_L-q_H}{q_L}-\tfrac{1}{p}\tfrac{q_Hq_L-q_M^2}{q_L(q_H-q_M)}-p\left(\tfrac{q_H}{q_M}-\tfrac{q_M}{q_L}\right)\right) < p\left(\tfrac{q_M}{q_L}-\tfrac{q_H}{q_L}\right)+\tfrac{q_M^2-q_Hq_L}{q_L(q_H-q_M)}.
\end{aligned}$$

Note that, for any $p > 0$, as $N \to \infty$, $p_L^{PAM} \to 0$. The condition therefore simplifies to,

$$0 < p < \frac{q_M^2 - q_Hq_L}{(q_H-q_M)^2} := \tilde{p} < 1,$$

where the inequality follows from the assumption of strict supermodularity. For any $\epsilon > 0$, we can therefore find a $\tilde{N}$ such that, for any $N > \tilde{N}$, if $p < \tilde{p} - \epsilon$, then expected wage payments under PAM exceed those under delegation. As both delegation and PAM implement the same matching, delegation therefore yields higher profits than PAM.

Now, we compare delegation to an arbitrary non-assortative matching with interim probability $p_H^\mu \in [p_H^{NAM}, p]$. By Equation 1 in the proof of Theorem 1, the loss in allocative efficiency arising from implementing such a matching is,

$$(p_H^{PAM} - p_H^\mu)\frac{p}{2}(q_H + q_L - 2q_M) > 0.$$

And using Equation 3, the difference in expected rent payments is,

$$\begin{aligned}
& cp\left[p_H^{PAM}\left(\tfrac{q_H+q_L-2q_M}{q_L}\right)+2\left(\tfrac{q_M}{q_L}-1\right)\right]-cp\left[p_H^\mu\tfrac{p}{1-p}\left(\tfrac{q_M}{q_L}-\tfrac{q_H}{q_M}\right)+\tfrac{p}{1-p}\left(\tfrac{q_H}{q_M}-\tfrac{q_M}{q_L}\right)+\tfrac{q_M}{q_L}-1\right]\\
= \; & cp\left[\tfrac{q_M}{q_L}-1-(1-p_H^\mu)\tfrac{p}{1-p}\left(\tfrac{q_H}{q_M}-\tfrac{q_M}{q_L}\right)+p_H^{PAM}\left(\tfrac{q_H+q_L-2q_M}{q_L}\right)\right] > 0,
\end{aligned}$$

since $p_L^\mu = \frac{p}{1-p}(1-p_H^\mu) < 1$. Hence, delegation yields higher profits if and only if,

$$c \leq \frac{1}{2}\left(\frac{(p_H^{PAM}-p_H^\mu)(q_H+q_L-2q_M)}{\frac{q_M}{q_L}-1-(1-p_H^\mu)\frac{p}{1-p}\left(\frac{q_H}{q_M}-\frac{q_M}{q_L}\right)+p_H^{PAM}\left(\frac{q_H+q_L-2q_M}{q_L}\right)}\right),$$

for all $p_H^\mu \in [p_H^{NAM}, p]$. The derivative of the right-hand side expression with respect to $p_H^\mu$ is proportional to a negative term,

$$-\left(\frac{q_M}{q_L}-1+\frac{p}{1-p}\frac{q_M^2-q_Hq_L}{q_Lq_M}(1-p_H^{PAM})+p_H^{PAM}\frac{q_H+q_L-2q_M}{q_L}\right)<0.$$

Hence, it is minimized when $p_H^\mu = p$. Consequently, delegation outperforms both NAM and RM if and only if,

$$0 < c \leq \frac{1}{2}\left(\frac{(p_H^{PAM}-p)(q_H+q_L-2q_M)}{\left(\frac{q_M}{q_L}-1\right)-p\left(\frac{q_H}{q_M}-\frac{q_M}{q_L}\right)+p_H^{PAM}\left(\frac{q_H+q_L-2q_M}{q_L}\right)}\right) := \tilde{c}.$$

# B Global Distortion of PAM

Throughout our analysis, we have assumed that the manager restricts attention to contracts that induce effort by all workers in all teams. Under strict log submodularity, we found conditions under which PAM is distorted given this restriction. We now find conditions under which distorting PAM is optimal even when consider the possibility of not implementing effort in some teams. We focus on the case in which the probability of highs is large and the firm is large.[28]

**Proposition 1.** *Suppose $N$ is large, $p \geq \frac{1}{2}$, $2q_M$ is close to $q_H+q_L$, and $2q_M > q_H$. Then, there exists $\hat{c} > \bar{c}$ such that if $c \in (\bar{c}, \hat{c})$, then implementing NAM and effort by all workers in all teams yields strictly higher profits than implementing PAM and any effort implementation. Hence, distorting PAM is globally optimal.*

[28] The large firm case is of special interest since the output distortion introduced by implementing NAM rather than PAM is increasing in $N$: $p_H^{PAM}$ is increasing in $N$ and $p_H^{NAM}$ is decreasing in $N$. A derivation of this result is available upon request.

*Proof.* As $N \to \infty$, $p_H^{PAM} \to 1$ and, if $p \geq \frac{1}{2}$, $p_H^{NAM} \to \frac{2p-1}{p}$.[29] Hence, under PAM, the expected number of teams with one high and one low converges to zero in the large $N$ limit and we need only consider three possibilities: (i) implementing effort only in teams composed of two lows; (ii) implementing effort only in teams composed of two lows and teams composed of two highs; and (iii) implementing effort only in teams composed of two highs. But, implementing effort only in teams composed of two lows is infeasible since highs would have an incentive to misreport their type, and implementing effort only in teams composed of two lows and teams composed of two highs is equivalent to implementing PAM and effort by all workers in all teams in the limit. Hence, if $c > \bar{c}$, so that implementing NAM and effort by all workers in all teams outperforms implementing PAM and effort by all workers in all teams, then it suffices to compare implementing NAM and effort by all workers to implementing PAM and high effort only in teams composed of two highs to establish a global distortion of PAM.

For any equal treatment matching, the optimal anonymous, independent wage scheme implementing effort by two highs in a team is $w_H^H(1) = \frac{c}{q_H}$ and all other wages equal to zero. Profit comparisons may be made by comparing expected output produced per worker net expected wage paid per worker. In

[29]The expressions for $p_H^{PAM}$ and $p_H^{NAM}$ are given by:

$$P_H^{PAM} = \sum_{k=0}^{N/2-1} \binom{N-1}{2k} (1-p)^{2k} p^{N-2k-1} + \sum_{k=0}^{N/2-1} \binom{N-1}{2k+1} (1-p)^{2k+1} p^{N-2k-2} \frac{N-2k-2}{N-2k-1}$$

$$P_H^{NAM} = 1 - \frac{1-p}{p} P_L^{NAM} = 1 - \frac{1-p}{p} \left( \sum_{k=0}^{N/2-1} \binom{N-1}{k} (1-p)^k p^{N-k-1} + \sum_{k=0}^{N/2-1} \binom{N-1}{k} (1-p)^{N-k-1} p^k \frac{k}{N-k} \right).$$

A derivation of the large $N$ limit result is available upon request.

the case of implementing PAM and effort by two highs, this simplifies to,

$$p\left[\frac{q_H}{2}-c\right].$$

If $p \geq \frac{1}{2}$, so that $p_H^{NAM}=\frac{2p-1}{p}$ in the large firm limit, the corresponding expression when implementing NAM and effort in all teams is

$$\frac{2p-1}{2}q_H+(1-p)q_M-cp[\frac{q_H}{q_M}-1]-c.$$

Hence, NAM and effort in all teams yields higher profits than PAM and effort only in teams composed of two highs if,

$$p\frac{q_H-2c}{2}<\frac{2p-1}{2}q_H+(1-p)q_M-cp[\frac{q_H}{q_M}-1]-c \Leftrightarrow c[p\frac{q_H}{q_M}-2p+1]<(1-p)q_M+\frac{p-1}{2}q_H.$$

Notice that when $2q_M > q_H$, the right-hand side is positive. Further, the left-hand side is positive for any $p$. It follows that NAM and effort in all teams yields higher profits than PAM and high effort in teams composed of two highs if

$$c<\frac{\frac{1-p}{2}(2q_M-q_H)}{p\frac{q_H}{q_M}-2p+1}:=\hat{c}.$$

It remains to check that $\bar{c}>\hat{c}$. This happens if,

$$\begin{aligned}
&\frac{\frac{1-p}{2}(2q_M-q_H)}{p\frac{q_H}{q_M}-2p+1}>\frac{1}{2}\left(\frac{1-p}{p}\right)\left(\frac{q_H+q_L-2q_M}{\frac{q_M}{q_L}-\frac{q_H}{q_M}}\right)\\
\Leftrightarrow\quad &2\frac{q_M^2}{q_L}-2q_H-\frac{q_Hq_M}{q_L}+\frac{q_H^2}{q_M}>\frac{q_H^2}{q_M}+\frac{q_Hq_L}{q_M}-2q_H+\left(\frac{1}{p}-2\right)(q_H+q_L-2q_M)\\
\Leftrightarrow\quad &2\frac{q_M^2}{q_L}-\frac{q_Hq_M}{q_L}-\frac{q_Hq_L}{q_M}>\left(\frac{1}{p}-2\right)(q_H+q_L-2q_M).
\end{aligned}$$

The right-hand is less than zero since $q_H+q_L-2q_M>0$ under strict supermodularity and $\frac{1}{p}-2\leq 0$ when $p\geq\frac{1}{2}$. The left-hand side is positive if $2q_M>q_H+q_L\frac{q_H}{q_M}\frac{q_L}{q_M}$, which holds when $2q_M\approx q_H+q_L$ since strict log submodularity implies that $\frac{q_H}{q_M}\frac{q_L}{q_M}<1$. □