Moral Hazard, Firm Size, and the Size-Wage Differential

Michael T. Rauh
Indiana University
mtrauh@indiana.edu

March 11, 2015

Abstract

Does moral hazard limit the size of the firm relative to the first best? We develop a teams model where team size is endogenous. We show that when individual output is contractible greater employment reduces the moral hazard problem and the second best firm can be larger or smaller than the first best firm. When only team output is contractible greater employment increases the moral hazard problem which limits the size of the firm. We consider the relationship between employment, optimal incentives, and wages and determine when the expected transfer is increasing in firm size consistent with the size-wage differential.

JEL Classifications: D02, D21, D86, L25, M5.

Keywords: division of labor, employment, endogenous team size, incentives, moral hazard, size-wage differential, size of the firm, specialization, teams, wages.
1 Introduction

A classic insight from contract theory is that moral hazard increases the cost of implementing high effort. If the principal (she) cannot observe the effort choice of the agent (he) then she must offer incentives to induce him to choose high effort. Incentives expose the agent to risk so if he is risk averse then additional compensation will be required to meet the participation constraint. But if moral hazard has implications for compensation it will also have implications for employment. In particular, there may be a relationship between the extent of moral hazard and the size of the firm as measured by employment. This also raises questions about the interrelations between employment, incentives, and wages. Are incentives and employment strategic complements or substitutes for the principal? What is the relationship between wages and firm size?

In this paper we develop a teams model in the spirit of Holmström (1982) where team size is endogenous and chosen by the principal. We consider a production technology similar to that in Becker and Murphy (1992) which involves a set of tasks that are equally divided among agents. Each agent produces output in each of his assigned tasks and these task outputs are then assembled into individual output where one unit of the latter requires one unit of the former in each assigned task. The individual outputs are then assembled into team output where again we make the Leontief assumption that it takes one unit of the former from each agent to produce one unit of the latter. Since we are interested in moral hazard we add a stochastic element to the deterministic Becker and Murphy model so the principal cannot perfectly infer the agents’ efforts. Specifically, there is some probability that individual output will be low (normalized to zero) even when the agent chooses high effort and this probability is increasing in the proportion of tasks that he performs. Assuming independence, the probability that team output is high is the product of the probabilities that the individual outputs are high. We formally introduce the model in section 2 below.

We assume this production technology for several reasons. First, it is analytically tractable and produces results which are directly comparable to the textbook moral hazard model with one agent. This allows us to clearly trace the effects of moral hazard on the principal’s employment decision and the corresponding feedback effects between employment, incentives, and wages. We also believe that it describes reasonably well many real-world production processes where production is divided into separate tasks assigned to different workers and the individual outputs are then aggregated into team output. In such settings a primary concern is the possibility of production failures that

---

1 This is referred to as the risk-reward tradeoff. For a survey on the (mixed) empirical evidence see Prendergast (1999).
2 See for example Laffont and Martimort (2002, Chapter 4).
result in delays or quality defects which require costly re-work. It is precisely this concern about
the possibility of delays which leads Ichmowski, Shaw, and Prennushi (1997) to take “uptime”
as the main performance measure in their empirical study of steel finishing lines. In our model
this is captured by the possibility of negative shocks to individual output which result in negative
shocks to team output because of the Leontief assumption. One interpretation is simply that with
some probability an agent makes an independent mistake and the probability of such mistakes
increases with the share of tasks performed by the agent. Another interpretation is that with some
probability a production failure occurs in at least one of the agent’s tasks. This results in quality
defects which render output effectively zero in all affected tasks. Each agent independently monitors
the production process and detects such failures with some probability which is decreasing in the
share of tasks performed by the agent. When problems are detected they are immediately fixed
but undetected problems result in zero individual and team output.

The employment decision is characterized by the following tradeoffs: an increase in employment
allows for greater specialization and division of labor (SDOL) which improves productivity; it
reduces the share of tasks performed by each agent and therefore raises the individual probability
(IP) that individual output is high; for given individual probabilities, an increase in the number
of agents increases the number of stages of production where a negative shock could occur which
reduces the team probability (TP) that team output is high; and greater employment can change
the extent of the moral hazard problem (MH) and hence the optimal incentive and wage bill of the
principal. As in Becker and Murphy, the SDOL effect generates increasing returns to employment.
Note that the combined effect of IP and TP on the probability that team output is high is ambiguous
because an increase in employment raises the probability that individual output is high but also
adds more probabilities to the product which, assuming independence, defines the probability that
team output is high.

To determine the effect of moral hazard on employment we compare the second best firm with
the first best firm where there is no moral hazard. We therefore begin in section 3 with the first best
benchmark. As usual when the agents’ efforts are contractible the principal provides full insurance
so the first best transfer does not depend on the probability that individual output is high. In our
model this standard result implies that the first best transfer does not depend on employment so
the first best cost of employment is therefore linear in employment. But a linear cost of employment
cannot constrain the increasing returns due to the SDOL effect and the first best firm will therefore
be infinite unless the negative TP effect on the probability that team output is high dominates
both SDOL and the IP effect on the probability that individual output is high.
In section 4 we introduce moral hazard. We first consider the case where individual output is contractible. When agents choose high effort an increase in employment has the IP effect of raising the probability that individual output is high. For a given level of incentives this increases the expected reward for high effort and reduces the moral hazard problem. The second best incentive and wage (i.e., expected transfer) therefore decline with employment and incentives and employment are strategic substitutes for the principal. In the limit the moral hazard problem vanishes and the second best incentive and wage converge to their first best levels. The second best cost of employment is therefore \textit{asymptotically} linear and first best. We derive several conclusions from this. As was the case for the first best firm, the second best firm is infinite unless TP dominates SDOL and IP. In particular, the second best firm is finite iff the first best firm is finite and when both are finite the second best firm can be larger or smaller than the first best firm.\footnote{As far as we know the only other paper where something similar occurs is Stole and Zwiebel (1996) who show that the principal employs more workers than the neoclassical firm to reduce the workers’ bargaining power.} The former occurs when greater employment reduces the moral hazard problem to such an extent that the second best marginal cost of employment is smaller than its first best counterpart. We conclude that when individual output is contractible the effect of moral hazard on employment is ambiguous.

The significance of these results is attested to by the stylized fact in the management literature [see Carter and Keon (1986) for a survey] that larger firms have more specialized workers and a more extensive division of labor. Furthermore, the result that the relationship between employment and moral hazard is ambiguous when individual performance is contractible seems robust to certain extensions of the model. For example, Liang, Rajan, and Ray (2008) consider a version of the linear model in Holmström and Milgrom (1987) where the variance of the additive normally distributed productivity shock is assumed to be increasing in employment. In that case greater employment makes performance harder to measure and exacerbates the moral hazard problem. But when a production process is characterized by SDOL one might imagine that an increase in specialization would make performance \textit{easier} to measure and this is indeed the case in our model. Likewise, Auriol, Friebel, and Pechlivanos (1999) consider a version of the linear model where an increase in employment reduces the positive effect on individual output of other agents’ helping efforts. In their model the second best firm is always smaller than the first best firm when agents are risk averse. But individual performance is actually non-contractible in their model because individual output depends on the effort choices of the entire team. Their paper is therefore an analysis of the teams problem to which we now turn.

In section 5 we consider the \textit{teams} problem where the only contractible performance measure is
team output. We first identify necessary and sufficient conditions such that TP dominates IP so the probability that team output is high declines with employment. In that case greater employment reduces the expected reward for high effort and exacerbates the moral hazard problem. As a result the optimal team incentive and wage increase with employment so incentives and employment are strategic complements for the principal. The prediction that wages increase with firm size has been repeatedly confirmed in the empirical labor economics literature where it is referred to as the size-wage differential. For example, Brown and Medoff (1989) show that differences in worker quality can explain about one-half of the size-wage differential but find little or no evidence in support of any of the other conventional explanations. In our model wages increase with firm size (measured in terms of employment) because the moral hazard problem becomes more acute.

Unlike the case where individual output is contractible, the optimal team incentive, wage, and marginal cost of employment do not converge to their first best levels but rather explode with greater employment. We provide sufficient conditions such that MH dominates SDOL so the optimal team size is finite. In this case the size of the firm is bounded by the growing moral hazard problem which makes the marginal cost of employment rise sufficiently rapidly relative to the increasing returns due to specialization and division of labor. Under fairly general conditions we show that the optimal team size is smaller than first best. Finally, we provide an analytically tractable example to illustrate our main results. We conclude that it is only in the teams setting that moral hazard limits the size of the firm.

Related literature includes Becker and Murphy (1992), who consider essentially the first best employment decision (no moral hazard). To bound the size of the firm they include an ad hoc cost of employment which they loosely justify in terms of coordination costs. To study the impact of moral hazard on employment one must have a model where team size is endogenous. As far as we know the only such papers are Auriol, Friebel, and Pechlivanos (1999); Liang, Rajan, and Ray (2008); and Rauh (2014). The latter paper extends the Becker-Murphy framework to the case of moral hazard where only team output is contractible. Like Becker and Murphy, the model includes an ad hoc employment cost which is unsatisfactory for the present purpose of exploring the relationship between moral hazard and firm size. All three papers assume linear contracts which is unsuitable for an explanation for the size-wage differential which requires a model where wages are set optimally.

Section 6 concludes. All proofs are in the appendix.
2 Model

We consider one principal and a population $\mathbb{N}$ of identical agents, where $\mathbb{N} = \{1, 2, \ldots\}$ is the set of positive integers. Let $n \geq 0$ be the number of agents employed by the principal. Following Becker and Murphy, the production process consists of a set $S = [0, 1]$ of tasks. Let $S_i \subseteq S$ be the set of tasks assigned to agent $i$. We assume an equal division of tasks in the sense that $\{S_i\}_{i=1}^n$ is a partition of $S$ such that each $S_i$ has measure $\rho = 1/n$ for all $i$.\footnote{An equal division of tasks would be optimal if each task requires an initial training period with the costs borne by the principal and these costs increase at an increasing rate for each agent. It would also be optimal if an unequal division of tasks results in time delays because agents with more tasks require more time to complete them. For simplicity we do not include these considerations in the model.}

After receiving his task assignment, for each task $s \in S_i$ agent $i$ chooses his investment $l_{is} \geq 0$ in task-specific human capital and production effort $e_{is} \geq 0$. Once these decisions are made by all agents and for all tasks an independent random shock is realized for each agent. Let $q_{is}$ denote output in task $s \in S_i$. With probability $\pi(\rho)$ the shock is positive and the task outputs are high $q_{is} = \overline{q}_{is}$ for all $s \in S_i$, where

$$\overline{q}_{is} = A l_{is}^{\gamma} e_{is}$$

and $A > 0$ and $\gamma > 0$ are productivity parameters. With probability $1 - \pi(\rho)$ the shock is negative and task output is low $q_{is} = 0$ for at least one task $s \in S_i$.

After all shocks are realized each agent $i$ assembles the task outputs $q_{is}$ into individual output $q_i$. The production technology is Leontief in the sense that one unit of individual output requires one unit of task output in each task:

$$q_i = \inf_{s \in S_i} q_{is}.$$ 

We make the following assumptions on $\pi(\rho)$.

**Assumption 1** $\pi : [0, 1] \to [0, 1]$ is (i) continuous, (ii) $\pi(\rho) > 0$ for all $\rho \in (0, 1]$ and $\pi(0) = 1$, (iii) twice continuously differentiable with $\pi' < 0$ on $(0, 1]$.

Apart from the technical requirements of continuity and differentiability, the main assumption is that the probability $\pi(\rho)$ of a positive shock declines as the share $\rho$ of tasks performed by the agent increases. This is the IP effect described in the introduction where we also provided two simple interpretations. As $\rho \to 0$ the probability of a positive shock converges to one. Note that we do not assume $\pi(1) = 0$ because we do not want to rule out a single-agent firm ex ante. In the appendix we show that the following example satisfies all of the above assumptions.
Example

Let \( \pi(\rho) = \left( \frac{\rho}{D} \right)^\rho \) and \( \pi(0) \equiv 1 \), where \( D > e \) and \( e \) is Euler’s constant. □

Since shocks are independent across agents the probability that all the individual outputs are high is given by \( p(n) = \pi(\rho)^n \). Once the individual outputs are determined they are then assembled into team output \( Q \). We again assume that one unit of team output requires one unit of individual output from each agent so that \( Q = \min_i q_i \).

We now characterize the population \( N \) of ex ante identical agents. Each employed agent \( i \) chooses total effort \( T_i \) which can be low or high \( T_i \in \{0, 1\} \). The agent must then allocate his total effort \( T_i \) across tasks in \( S_i \). Let \( \tau_{is} \) be the total time that \( i \) devotes to task \( s \in S_i \), where

\[
T_i = \int_{s \in S_i} \tau_{is} \, ds. \tag{3}
\]

The choice of production effort \( e_{is} \) and investment \( l_{is} \) in task-specific human capital must therefore satisfy \( e_{is} \geq 0 \), \( l_{is} \geq 0 \), and

\[
e_{is} + l_{is} \leq \tau_{is}. \tag{4}
\]

We assume each agent has an outside option of zero and that the investments \( l_{is} \) are firm-specific as well as task-specific so the outside option is unaffected. All agents have the same utility function

\[
U(t_i, T_i) = u(t_i) - c(T_i), \tag{5}
\]

where \( u \) is the utility of money \( t_i \) and the cost \( c \) of total effort \( T_i \) is given by \( c(0) = 0 \) and \( c(1) = \psi > 0 \). Let \( \mathbb{R} \) be the set of real numbers, \( \mathbb{R}_+ \) the nonnegative real numbers, and \( \mathbb{R}_++ \) the positive real numbers.

**Assumption 2** \( u : \mathbb{R}_+ \to \mathbb{R} \) is (i) continuous, (ii) twice continuously differentiable on \( \mathbb{R}_++ \) with \( u' > 0 \) and \( u'' < 0 \), and (iii) \( u(0) = 0 \). Let \( t = h(u) \) be defined by \( h = u^{-1} \) and \( \epsilon_h = uh'/h \) be the elasticity of \( h \) with respect to \( u \). (iv) \( \epsilon_h \) is bounded on \( \mathbb{R}_++ \) and (v) \( h' \to \infty \) as \( u \to \infty \).

The utility function \( u \) is only defined for nonnegative transfers \( t_i \) so the principal cannot fine the agent. This seems realistic but does not appear to play a significant role in our results. Otherwise, requirements (i)-(iii) are fairly standard assumptions to the effect that agents are risk averse. The assumption (iv) that the elasticity \( \epsilon_h \) of \( h \) is bounded is a mild technical requirement whose role

---

\(^5\)We implicitly assume that \( \tau_{is} \) is a measurable function on \( S_i \) such that \( 0 \leq \tau_{is} \leq 1 \) for all \( s \in S_i \) and \( i \).
will be made explicit later. Condition (v) states that the marginal utility $u'$ of money converges to zero as $t_i \to \infty$. An example which satisfies all of these assumptions is $u(t_i) = \sqrt{t_i}$. As the name suggests the following example will be used repeatedly throughout the paper.

**Main Example**

Let $u(t_i) = \sqrt{t_i}$ and $\pi(\rho) = (\rho/D)^\rho$, where $\pi(0) \equiv 1$ and $D > e$. □

3 The First Best

To gauge the effects of moral hazard on the employment decision we first consider the first best problem where there is no moral hazard and all the agents’ choices are contractible. Let $t_i$ be the transfer to agent $i$ when individual output is high $q_i = \overline{q}_i$ and $t_i$ the transfer when $q_i = 0$. The principal chooses employment $n \geq 0$, production efforts $\{e_{is}\}$, investments in human capital $\{l_{is}\}$, total task efforts $\{\tau_{is}\}$, total efforts $\{T_i\}$, and transfers $\{\vec{t}_i\}$ and $\{\vec{t}_i\}$ to maximize expected profit

$$\Pi^F = p(n)Q - \sum_{i=1}^{n} \left\{ \pi(\rho)\vec{t}_i + [1 - \pi(\rho)]\vec{t}_i \right\}$$

subject to (3), (4), $T_i \in \{0, 1\}$ for all $i$, and the relevant participation constraints discussed later. The first term is expected output or revenue. In particular, team output is high $Q = \min_i q_i$ when all the individual outputs are high which occurs with probability $p(n)$. With probability $1 - p(n)$ at least one individual output is zero and team output is zero. The second term is the sum of the expected transfers to agents. If $n = 0$ then $\Pi^F \equiv 0$.

Given that agent $i$ chooses high effort $T_i = 1$, our first result shows how it should be allocated across tasks and how total effort $\tau_{is}$ in task $s \in S_i$ should be allocated between production effort $e_{is}$ and investments $l_{is}$ in human capital.

**Proposition 1** If the principal chooses high total effort $T_i = 1$ for agent $i$ then the allocation of $T_i = 1$ that maximizes high individual output $\overline{q}_i$ in (2) is given by

$$e = \frac{1}{\rho(1 + \gamma)}, \quad l = \frac{\gamma}{\rho(1 + \gamma)}, \quad \text{and} \quad \tau = \frac{1}{\rho},$$

where we have eliminated the unnecessary subscripts. In that case, high individual and team output

---

6 We have normalized the price $r$ of output to be one. Otherwise, the term $rB$ would appear everywhere that $B$ does so the price can be included in the parameter $B$. 
are given by
\[ \bar{q} = \bar{Q} = Bn^{1+\gamma}, \]  
\[ (8) \]
where
\[ B = \frac{A}{\gamma} \left( \frac{\gamma}{1+\gamma} \right)^{1+\gamma}. \]
\[ (9) \]

Since individual output is essentially the minimum of the corresponding task outputs it is optimal to allocate total effort \( T_i = 1 \) equally \( \tau_{is} = 1/\rho \) across all tasks \( s \in S_i \). One can then verify that the choice of \( e_{is} \) and \( l_{is} \) which maximizes high task output \( \bar{q}_{is} \) in (1) subject to the constraints \( e_{is} \geq 0, l_{is} \geq 0, \) and (4) is given by (7). Under an equal division of labor \( \rho = 1/n \) an increase in employment corresponds to greater specialization and division of labor which allows each agent to devote more time \( \tau = 1/\rho = n \) to each task \( s \in S_i \). This results in greater production effort \( e \) and investments \( l \) in human capital in each task and hence higher individual output \( \bar{q} \). If \( \gamma > 0 \) then high team output per worker \( \bar{Q}/n \) is increasing in employment which is characteristic of the increasing returns traditionally associated with specialization and division of labor. This is the SDOL effect described in the introduction.

**Definition 1** We say the principal implements high efficient effort if \( T_i = 1 \) for all \( i \) and \( T_i = 1 \) is allocated as in Proposition 1 for all \( i \).

We now list the options available to the principal for a given set of transfers:

1. zero employment \( n = 0 \) which gives zero profit \( \Pi^F = 0 \),
2. \( n \geq 1, T_i = 0 \) for at least one \( i \) (so team output is zero for sure), and \( \Pi^F \leq 0 \),
3. \( n \geq 1, T_i = 1 \) for all \( i \), and an inefficient allocation of \( T_i = 1 \) for at least one \( i \), and
4. \( n \geq 1 \) and high efficient effort.

We assume the principal prefers \( n = 0 \) over \( n \geq 1 \) and \( T_i = 0 \) for all \( i \),\(^8\) so the principal’s choice boils down to options (1) and (4). For the rest of the paper we assume that in all contexts the

\(^7\)We do not need precise definitions in this paper but for completeness the degree of specialization of agent \( i \) is given by \( (1/\rho_i) - 1 \); i.e., a more specialized agent performs fewer tasks. The extent of the division of labor is given by the inverse of the Herfindahl index \( \sum_{i=1}^{n} \rho_i^2 \) which is maximized under an equal division of labor.\(^8\)This would be the case if agents had a positive instead of zero outside option.
principal chooses the latter.\(^9\) Under this assumption

\[
\Pi^F = p(n)Bn^{1+\gamma} - \sum_i \{\pi(\rho)|\bar{t}_i + [1 - \pi(\rho)]|L_i\} 
\]  

(10)

with participation constraints

\[
\pi(\rho)u(\bar{t}_i) + [1 - \pi(\rho)]u(L_i) - \psi \geq 0.
\]

(11)

**Proposition 2** The first best solution entails a constant transfer \(\bar{t}_i = L_i = h(\psi)\) for all \(i\).

As usual the first best transfers provide full insurance \(\bar{t}_i = L_i = h(\psi)\) which is the same constant transfer as in the textbook moral hazard model with one agent. In our model this standard result implies that the first best cost \(C^F(n, \psi) = nh(\psi)\) of employment is linear in \(n\). Substituting these transfers into (10),\(^10\)

\[
\Pi^F(n | B, \psi) = p(n)Bn^{1+\gamma} - C^F(n, \psi).
\]

(12)

We are now ready to study the first best employment decision. Let \(n^F(B, \psi)\) be the first best employment correspondence.\(^11\) Note that we define \(n^F\) in terms of \(\psi^{-1} = 1/\psi\) rather than \(\psi\) for reasons that will become clear shortly. Let \(\emptyset\) denote the empty set.

**Proposition 3**

(i) If \(p(n)n^\gamma \to 0\) as \(n \to \infty\) then \(n^F(B, \psi)\) \(\neq \emptyset\) and every \(n \in n^F(B, \psi)\) is finite for all \(B > 0\) and \(\psi > 0\).

(ii) If \(p(n)n^\gamma \to \infty\) as \(n \to \infty\) then the first best employment level is infinite for all \(B > 0\) and \(\psi > 0\).

The size of the first best firm is determined by the following factors. First, the increasing returns to employment due to SDOL evident in the term \(\overline{Q} = Bn^{1+\gamma}\) Second, the direct cost \(C^F\) of employment whose effect is limited because it is only linear in \(n\). Third and finally, the effect of employment on the probability \(p(n) = \pi(\rho)^n\) that team output is high \(\overline{Q}\). An increase in employment has the IP effect of raising the probability \(\pi(\rho)\) that individual output is high but it

\(^9\) A simple sufficient condition for the first best problem is \(\pi(1)B > h(\psi)\) which ensures positive expected profit from hiring at least one agent.

\(^10\) We do not include \(\gamma > 0\) as an argument in \(\Pi^F\) because it is held constant throughout the paper except in examples.

\(^11\) When employment \(n \geq 0\) must be an integer one cannot make any reasonable assumptions to ensure uniqueness.
also has the \textit{TP effect} of increasing the number of stages of production at which a negative shock could occur. If the TP effect dominates so that \( p(n) \to 0 \) sufficiently rapidly then \( p(n)n^\gamma \to 0 \) as \( n \to \infty \) and the first best firm will be finite. But if the IP and SDOL effects dominate and \( p(n) \to 0 \) too slowly or is bounded away from zero for sufficiently large \( n \) then \( p(n)n^\gamma \to \infty \) as \( n \to \infty \) and the first best firm will be infinite.\footnote{Another possibility is that \( p(n)n^\gamma \) converges to a positive constant. In that case, the first best employment level is zero, indeterminate, or infinite depending on the parameters; see our Main Example with \( \gamma = 1 \) and the proof of Proposition 3. We do not consider cases where \( p(n)n^\gamma \) fails to converge.}

\textit{Main Example}

If \( \pi(\rho) = (\rho/D)^{\rho} \) and \( D > e \) then \( p(n)n^\gamma = (1/D)n^{\gamma - 1} \). If \( 0 \leq \gamma < 1 \) then \( p(n)n^\gamma \to 0 \) and the first best firm is finite according to the above result. If \( \gamma > 1 \) then \( p(n)n^\gamma \to \infty \) and the first best firm is infinite. Indeed, the expected first best profit is

\[ \Pi^F = \frac{B}{D} n^\gamma - n\psi^2 \]  

(13)

and for all \( 0 \leq \gamma < 1 \) the first best employment level is

\[ n^F = \left( \frac{B\gamma}{D\psi^2} \right)^{\frac{1}{1-\gamma}} \]  

(14)

ignoring the integer constraint. If \( \gamma = 1 \) then

\[ n^F = \begin{cases} 
0 & \text{if } B/D < \psi^2 \\
\text{indeterminate} & \text{if } B/D = \psi^2 \\
\infty & \text{if } B/D > \psi^2.
\end{cases} \]  

(15)

If \( \gamma > 1 \) then \( n^F = \infty \). □

The comparative statics of the first best solution are fairly straightforward once the appropriate definitions are made. Let \( \Omega^F \) be the set of all \( (B,\psi^{-1}) \in \mathbb{R}^2_{++} \) such that \( 0 \notin n^F(B,\psi^{-1}) \).\footnote{Footnote 9 indicates one way that \( \Omega^F \) could be constructed.} From
Proposition 1 The optimal correspondences for the remaining endogenous variables are:

\[ \rho^F(B, \psi^{-1}) = \left\{ \frac{1}{n} \mid n \in n^F(B, \psi^{-1}) \right\} \]  
(16)

\[ e^F(B, \psi^{-1}) = \left\{ \frac{n}{1 + \gamma} \mid n \in n^F(B, \psi^{-1}) \right\} \]  
(17)

\[ l^F(B, \psi^{-1}) = \left\{ \frac{n\gamma}{1 + \gamma} \mid n \in n^F(B, \psi^{-1}) \right\} \]  
(18)

\[ \tau^F(B, \psi^{-1}) = \left\{ n \mid n \in n^F(B, \psi^{-1}) \right\} \]  
(19)

\[ \pi^F(B, \psi^{-1}) = \left\{ \pi(1/n) \mid n \in n^F(B, \psi^{-1}) \right\} \]  
(20)

\[ Q^F(B, \psi^{-1}) = q^F(B, \psi^{-1}) = \left\{ Bn^{1 + \gamma} \mid n \in n^F(B, \psi^{-1}) \right\}. \]  
(21)

Proposition 4 Assume \( p(n)n^\gamma \to 0 \) as \( n \to \infty \).

(i) The first best employment correspondence \( n^F : \Omega^F \to \mathbb{N} \) is strongly increasing.\(^{14}\)

(ii) The first best proportion \( \rho^F \) of tasks is strongly decreasing on \( \Omega^F \).

(iii) The first best probability \( \tau^F \) of high individual output, production effort \( e^F \), investments \( l^F \) in human capital, total effort \( \tau^F \) per task, and high individual \( q^F \) and team \( Q^F \) output are all strongly increasing on \( \Omega^F \).

The comparative statics for (16)-(21) follow from those for employment \( n^F \), which is strongly increasing in the productivity parameter \( B \) and strongly decreasing in the cost of effort \( \psi \). In a first best world, firms with superior production technologies (higher \( B \)) or more motivated workers (lower \( \psi \)) are larger (higher employment \( n \)), have more specialized workers and a more extensive division of labor, a lower probability of production failures, greater effort and investments in human capital per task, and higher expected individual and team output.

4 Individual Output Contractible

We now introduce moral hazard. We assume that individual output \( q_i \) is contractible for all \( i \) but the principal cannot observe any of the agents’ choices \( \{e_{is}\}, \{l_{is}\}, \{\tau_{is}\}, \text{or} \{T_i\} \). Let \( t_i(q_i) \) be the transfer to agent \( i \) conditional on \( q_i \).\(^{15}\) Note that the principal does not condition transfers

\(^{14}\)Let \( T \) and \( S \) be partially ordered sets. A correspondence \( \phi : T \to S \) is strongly increasing if \( t \geq t' \) and \( t \neq t' \) imply \( s \geq s' \) for all \( s \in \phi(t) \) and \( s' \in \phi(t') \). Intuitively, an increase in the parameter \( t \) leads to an upward shift in the corresponding set \( \phi(t) \) of \( s \) values. The definition for strongly decreasing is similar.

\(^{15}\)In general, the transfer \( t_i(q_i) \) will also depend on other endogenous variables such as employment \( n \) as well as exogenous parameters such as the cost \( \psi \) of effort. To simplify notation we suppress all arguments in \( t_i \) except \( q_i \).
on team output $Q$ because that would impose additional risk on agents without adding any new information.

As in the previous section, the principal effectively chooses between zero employment and positive employment with high efficient effort. A principal who implements the latter chooses employment $n \in \mathbb{N}$ and transfers $t_i(q_i)$ to maximize expected profit

$$\Pi^F = p(n)\overline{Q} - \sum_{i=1}^{n} \{ \pi(\rho)t_i(\overline{\mathbb{Q}}) + [1 - \pi(\rho)]t_i(0) \}, \quad (22)$$

where $\overline{\mathbb{Q}} = \overline{q} = Bn^{1+\gamma}$, subject to $t_i(q_i) \geq 0$ for all $0 \leq q_i \leq \overline{q}$, the participation

$$\pi(\rho)u(t_i(\overline{\mathbb{Q}})) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq 0, \quad (23)$$

and relevant incentive compatibility constraints. To determine the latter we list the options available to agent $i$. If $i$ chooses $T_i = 0$ then $q_i = 0$ for sure. If $i$ chooses high efficient effort then

$$q_i = \begin{cases} \overline{q} & \text{with probability } \pi(\rho) \\ 0 & \text{with probability } 1 - \pi(\rho). \end{cases} \quad (24)$$

If $i$ chooses high inefficient effort then

$$q_i = \begin{cases} \hat{q}_i & \text{with probability } \pi(\rho) \\ 0 & \text{with probability } 1 - \pi(\rho), \end{cases} \quad (25)$$

where $0 \leq \hat{q}_i < \overline{q}$. The incentive compatibility constraints are therefore

$$\pi(\rho)u(t_i(\overline{\mathbb{Q}})) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq u(t_i(0)) \quad (26)$$

$$\pi(\rho)u(t_i(\overline{\mathbb{Q}})) + [1 - \pi(\rho)]u(t_i(0)) - \psi \geq \pi(\rho)u(t_i(\hat{q}_i)) + [1 - \pi(\rho)]u(t_i(0)) - \psi \quad (27)$$

for all $0 \leq \hat{q}_i < \overline{q}$. Constraint (26) states that agent $i$ prefers high efficient effort over $T_i = 0$ while (27) states that $i$ prefers high efficient effort over high inefficient effort.

**Proposition 5**

(i) For any employment level $n \geq 1$ the transfers

$$t_i(q_i) = \begin{cases} h \left[ \frac{\psi}{\pi(\rho)} \right] & \text{if } q_i = \overline{q} \\ 0 & \text{if } 0 \leq q_i < \overline{q} \end{cases} \quad (28)$$
minimize the expected cost of implementing high efficient effort.

(ii) The incentive $h(\psi/\pi)$ is increasing in the proportion $\rho$ of tasks and decreasing in $n$.

(iii) $h(\psi/\pi) \to h(\psi)$ as $n \to \infty$.

The first result establishes the optimal transfer for any given employment level $n \geq 1$ (optimal or not). By inspection the constraints (27) are satisfied iff $t_i(\hat{q}_i) \leq t_i(\bar{q})$ for all $0 \leq \hat{q}_i < \bar{q}$. For simplicity we focus on the optimal contract with the maximum punishment $t_i(\hat{q}_i) = 0$ for all $0 < \hat{q}_i < \bar{q}$. Note that this does not impose any additional risk because $0 < \hat{q}_i < \bar{q}$ occurs with zero probability when $i$ chooses high efficient effort. We now consider the two remaining constraints (23) and (26). Since increases in $t_i(0)$ merely increase what $t_i(\bar{q})$ has to be in order to satisfy (23) and (26), the principal sets $t_i(0) = 0$ and chooses the minimum $t_i(\bar{q}) = h(\psi/\pi)$ consistent with (23) and (26) such that both constraints bind.

The incentive $h(\psi/\pi)$ is the same as that in the textbook moral hazard model with one agent\textsuperscript{16} except that here the probability $\pi$ of high individual output is a function of the proportion $\rho$ of tasks assigned to $i$ and hence a function of employment $n$. In particular, an increase in employment has the IP effect of raising the probability that individual output will be high. This increases the expected reward from high efficient effort so each agent can be provided with weaker incentives. When individual output is contractible the model therefore predicts that employment and incentives are strategic substitutes for the principal and that larger firms (in the sense of employment) will have more specialized workers who receive weaker incentives. Some indirect evidence that incentives are weaker in larger firms is provided by Brown and Medoff (1989, p. 1054) and Rasmusen and Zenger (1990). In the limit the proportion of tasks performed by each agent becomes negligible, the probability of high individual output approaches one, and the second best incentive $h(\psi/\pi)$ converges to the first best transfer $h(\psi)$.

**Proposition 6**

(i) The second best expected transfer

$$H^I(\rho, \psi) = \pi(\rho)h\left[\frac{\psi}{\pi(\rho)}\right] \quad (29)$$

strictly exceeds the first best transfer $h(\psi)$ for all $0 < \rho \leq 1$ and $\psi > 0$.

\textsuperscript{16}C.f. equation (4.32) in Laffont and Martimort (2002, p. 160) with $\pi_0 = 0$. 
(ii) $H_I^I > 0$ for all $\psi > 0$, where the subscript indicates partial differentiation.\footnote{In examples we often ignore the integer constraint for employment for simplicity and in formal propositions where we take derivatives it should be understood that we are allowing $n$ or $\rho$ to be a continuous variable. But in all formal propositions an optimal employment level in all contexts must be a nonnegative integer.}

(iii) As $n \to \infty$

$$H^I(\rho, \psi) \to \pi(0)h\left[\frac{\psi}{\pi(0)}\right] = h(\psi).$$

The expected second best transfer $H^I = \pi h(\psi/\pi)$ exceeds the first best transfer for the usual reason that under moral hazard agents must be incentivized to supply high efficient effort but incentives impose risk which requires additional compensation. A reduction in the proportion of tasks performed by each agent due to an increase in employment raises the probability $\pi$ that individual output will be high but also reduces the incentive $h(\psi/\pi)$ as shown in the previous result. Since agents are risk averse, $h$ is strictly convex and the latter effect dominates. The moral hazard problem therefore declines with employment as does the second best incentive and expected transfer. In the limit the moral hazard problem vanishes altogether and the second best incentive and expected transfer converge to their first best values.

Substituting the second best transfers into (22) we obtain

$$\Pi^I(n | B, \psi) = p(n)Bn^{1+\gamma} - C^I(n, \psi),$$

where $C^I(n, \psi) = nH^I(\rho, \psi)$.

**Proposition 7** Let

$$\Delta^I(n, \psi) = 1 - \frac{\rho \pi'}{\pi} \left[1 - \frac{\psi h'(\psi/\pi)}{\pi h(\psi/\pi)}\right] = 1 - \epsilon_\pi(1 - \epsilon_h),$$

where $\epsilon_\pi$ is the elasticity of $\pi$ with respect to $\rho$ and $\epsilon_h$ is evaluated at $u = \psi/\pi$.

(i) The second best marginal cost of employment is given by

$$C_n^I = H^I(n, \psi)\Delta^I(n, \psi).$$

(ii) If $\epsilon_\pi \to 0$ as $\rho \to 0$ then $\Delta^I(n, \psi) \to 1$ as $n \to \infty$ for all $\psi > 0$ and

$$\lim_{n \to \infty} C_n^I(n, \psi) = C_n^F = h(\psi).$$
The first result shows that the second best marginal cost of employment can be decomposed into two components, $H^I$ and $\Delta^I$. The first component is that each additional hire must be paid the expected transfer $H^I$ while the second $\Delta^I$ captures the MH effect discussed in the introduction. In particular, each additional hire reduces the moral hazard problem and therefore reduces the expected second best transfer $H^I$. Indeed, $\Delta^I < 1$ because $\epsilon = \frac{\alpha'}{\pi}$ and in the appendix we show that $1 - \epsilon_h < 0$ because $h$ is strictly convex. The second result states that if the elasticity $\epsilon = \rho \pi'$ of $\pi$ with respect to $\rho$ converges to zero then the second best marginal cost of employment converges to its first best counterpart. The assumption $\epsilon \to 0$ may appear strong but since $\rho \to 0$ and $\pi \to 1$ a simple sufficient condition is that $\pi'$ is bounded on $(0, 1]$ or alternatively that $\pi' \to -\infty$ sufficiently slowly which are weak technical requirements.

Main Example

In our Main Example

$$\epsilon_n = \rho \left(1 + \log \frac{\rho}{\beta}\right) \to 0 \quad (35)$$

as $\rho \to 0$ because $\rho \log \frac{\rho}{\beta} \to 0$ (see the appendix). We can therefore apply the above result. But we can also verify it directly:

$$H^I = \psi^2 \left(\frac{D}{\rho}\right)^\rho \quad \Delta^I = 1 + \rho \left(1 + \log \frac{\rho}{\beta}\right)$$

$$C^I_n = \psi^2 \left[1 + \rho \left(1 + \log \frac{\rho}{\beta}\right)\right] \quad C^I_{\psi n} = 2 \psi \left(\frac{D}{\rho}\right)^\rho \left[1 + \rho \left(1 + \log \frac{\rho}{\beta}\right)\right]. \quad (36)$$

Here the decomposition $C^I_n = H^I \Delta^I$ is evident by inspection. Since $(D/\rho)^\rho = 1/\pi \to 1$ as $\rho \to 0$ we have $H^I \to \psi^2 = h(\psi)$ and $\Delta^I \to 1$. □

We now identify conditions under which the second best firm is finite. Let $n^I(B, \psi^{-1})$ be the second best employment correspondence. We assume the following for the rest of the paper.

Assumption 3 Assume $p(n)n^\gamma \to 0$ or $p(n)n^\gamma \to \infty$ as $n \to \infty$.

Note that this holds in our Main Example except for the case $\gamma = 1$.

Proposition 8

(i) If $p(n)n^\gamma \to 0$ then $n^I(B, \psi^{-1}) \neq \emptyset$ and every $n \in n^I(B, \psi^{-1})$ is finite.

(ii) If $p(n)n^\gamma \to \infty$ the second best employment level is infinite.

(iii) The second best firm is finite iff the first best firm is finite.
The clearest sense in which moral hazard could limit the size of the firm would be if the first best firm was infinite and the second best firm was finite. The above result shows that this cannot happen. From the previous result, the second best marginal cost of employment is asymptotically linear and the same as the first best. It follows that the size of both the first and second best firms is determined by \( p(n)n^\gamma \). In particular, both are finite if the latter converges to zero and both are infinite if it diverges. The intuition is the same as in the previous section.

A less drastic sense in which moral hazard could limit the size of the firm would be if the first and second best firms were both finite and the second best firm was smaller. The following example shows that when the second component \( \Delta^I \) dominates the first \( H^I \) the second best marginal cost of employment can be less than first best. In fact, \( \Delta^I \) and the second best marginal cost of employment can even be negative. As a result, the second best firm can be larger than the first best firm.

**Example**

Consider our Main Example with the specific values \( B = 7, \gamma = 1/2, D = 100, \) and \( \psi = 1/10 \). In Figure 1 below we plot \( C_n^I \) in (36) and \( C_n^F = \psi^2 \) as functions of employment \( n \).

![Figure 1](image_url)

Although not shown in the Figure, \( \Delta^I < 0 \) and therefore \( C_n^I < 0 \) from \( n = 1 \) until \( n \approx 5.27 \). For example, one can verify that \( C^I = 1 \) when \( n = 1 \) and \( C^I \approx 0.28 \) when \( n = 2 \). The two marginal cost curves cross at \( n \approx 29.09 \). After that, \( C_n^I \) gradually asymptotes to \( C_n^F \) as stated in Proposition 7. In Figure 2 below we plot \( \Pi^F \) and \( \Pi^I \) as functions of employment.
Not surprisingly, expected first best profit exceeds expected second best profit at every employment level. Note that \( \Pi^I < 0 \) for all \( n \geq 28 \) so \( C^I_n < C^F_n \) for all employment levels where expected second best profit is nonnegative. We would therefore expect that \( n^I > n^F \) because the second best firm has a smaller marginal cost of employment on the relevant region. In fact, \( n^I = 14 \) and \( n^F = 12 \) imposing the integer constraint. □

The second best firm can also be at least weakly smaller than the first best firm. Since both problems can have multiple solutions in the following result we consider the largest first best employment level \( n^F(B, \psi) = \max_n n^F(B, \psi^{-1}) \).

**Proposition 9** Let \( N^I(B, \psi) \) be the set of all integers \( n \geq 0 \) such that \( \Pi^I(n \mid B, \psi) \geq 0 \). If \( C^I_n(n, \psi) \geq h(\psi) \) for all \( n \in N^I(B, \psi) \) then \( n \leq \max_n n^F(B, \psi^{-1}) \) for all \( n \in n^I(B, \psi^{-1}) \).

If the second best marginal cost of employment is weakly greater than its first best counterpart for all employment levels where second best expected profit is nonnegative then every second best employment level will be weakly less than \( \pi^F(B, \psi^{-1}) \). In particular, if the first best problem has a unique solution then under the same circumstances every second best employment level will be weakly smaller than the first best employment level. From Propositions 8 and 9 and the previous example we conclude that when individual output is contractible the relationship between moral hazard and the size of the second best firm is ambiguous.

For the first best problem we obtained monotone comparative statics in part because the cross-partial \( C^F_{n\psi} = h'(\psi) > 0 \). In contrast the comparative statics of the second best problem can be non-monotonic. In the previous example we showed that the second best marginal cost of employment is negative when greater employment reduces the moral hazard problem to such an extent that \( \Delta^I < 0 \). From (36) we observe that \( C^I_{n\psi} \) has the same sign as \( C^I_n \) so the second best employment correspondence \( n^I \) is strongly increasing in \( \psi \) when both are negative. On this region
an increase in $\psi$ accentuates the moral hazard problem which the principal counters with an increase in employment. On the other hand, $\Delta^I > 0$ when the beneficial effect of employment on the moral hazard problem is sufficiently weak and if $C^I_n$ and $C^I_{n\psi}$ are globally positive then monotonicity is restored. Let $\Omega^I$ be the set of all $(B, \psi^{-1}) \in \mathbb{R}_+^2$ such that $0 \not\in n^I(B, \psi^{-1})$. Let $\rho^I$, $\pi^I$, $e^I$, $l^I$, $\tau^I$, $\bar{q}^I$, and $\bar{Q}^I$ be defined as in (16)-(21) except that $n^F$ is replaced by $n^I$.

**Proposition 10** Assume $p(n)n^{\gamma} \to 0$ as $n \to \infty$ and $C^I_n > 0$ and $C^I_{n\psi} > 0$ for all $n \geq 1$ and $\psi > 0$.

(i) The second best employment correspondence $n^I : \Omega^I \to \mathbb{N}$ is strongly increasing.

(ii) The second best proportion $\rho^I$ of tasks is strongly decreasing on $\Omega^I$.

(iii) The second best probability $\pi^I$ of high individual output, production effort $e^I$, investments $l^I$ in human capital, total effort $\tau^I$ per task, and high individual $\bar{q}^I$ and team $\bar{Q}^I$ output are all strongly increasing on $\Omega^I$.

**Example**

Let $D = 3$ in our Main Example. Although we cannot explicitly solve this problem we can still obtain some information about the solution using the above propositions. In particular, a plot of the expression $(D/\rho)^\rho \Delta^I$ in (36) with $D = 3$ reveals that it exceeds one for all $0 < \rho \leq 1$. It follows that $C^I_n > C^F_n > 0$ and $C^I_{n\psi} > 0$ for all $n \geq 1$ and $\psi > 0$ and the hypotheses of Propositions 9 and 10 hold in this case. If $0 \leq \gamma < 1$ and the productivity parameter $B$ is sufficiently large then the first best employment level (14) is unique, $n^F > 1$, and every second best employment level is at least weakly smaller with monotone comparative statics. □

5 The Teams Problem

We now consider the teams problem in the sense of Holmström (1982) where individual outputs $q_i$ are non-contractible and the only available contractible performance measure is team output $Q$.

We first characterize the behavior of the probability $p(n)$ that team output is high as a function of employment $n$.

**Proposition 11**

(i) $\epsilon_\pi > \log \pi(\rho)$ for all $\rho \in (0,1]$ $\iff$ $p'(n) < 0$ for all $n \geq 1$. 

18
(ii) $\pi' \to -\infty$ as $\rho \to 0 \iff p(n) \to 0$ as $n \to \infty$.

In the previous section the moral hazard problem declined with employment. It is only in the present teams context that we can formalize the intuitive idea that the moral hazard problem grows with employment. The first result is that the probability that team output is high is decreasing in employment iff the probability that individual output is high is relatively inelastic in the sense that $\epsilon_\pi > \log \pi(\rho)$. This ensures that the TP effect dominates the IP effect. Intuitively, $p(n) = \pi(\rho)^n$ converges to zero as $n \to \infty$ when the exponent $n$ compounds the fraction $\pi(\rho)$ to zero faster than $\pi(\rho)$ converges to one. In other words, $\pi(\rho)$ has to remain bounded away from one “long enough” for the exponent to drive the expression to zero. Since $\pi(\rho)$ converges to one as $\rho \to 0$, this convergence has to happen “at the last moment” and very rapidly. The condition $\pi' \to -\infty$ as $\rho \to 0$ ensures this. Note that both conditions are necessary and sufficient. In the appendix we show that our Main Example satisfies both of them.

**Assumption 4** For the rest of the paper we assume the hypotheses (or equivalently the conclusions) of Proposition 11.

A principal who chooses positive employment and implements high efficient effort chooses $n \in \mathbb{N}$ and transfers $t_i(Q)$ to maximize expected profit

$$p(n)Q - \sum_{i=1}^{n} \{p(n)t_i(Q) + [1 - p(n)]t_i(0)\}$$

subject to $t_i(Q) \geq 0$ for all $0 \leq Q \leq \overline{Q}$, the participation

$$p(n)u(t_i(\overline{Q})) + [1 - p(n)]u(t_i(0)) - \psi \geq 0$$

and incentive compatibility constraints

$$p(n)u(t_i(\overline{Q})) + [1 - p(n)]u(t_i(0)) - \psi \geq u(t_i(0))$$

$$p(n)u(t_i(\hat{Q})) + [1 - p(n)]u(t_i(0)) - \psi \geq p(n)u(t_i(\overline{Q})) + [1 - p(n)]u(t_i(0)) - \psi$$

for all $i$ and $0 \leq \hat{Q} < \overline{Q}$, where $\overline{Q} = Bn^{1+\gamma}$ is high team output. The constraint (39) states that agent $i$ prefers high efficient effort over zero effort while (40) states that $i$ prefers high efficient effort over high inefficient effort. In both cases this preference is conditional on all other agents choosing high efficient effort. The proof of the following result is similar to that for the analogous result in Proposition 5 and is omitted.
Proposition 12

(i) For any $n \geq 1$ the transfers that minimize the expected cost of implementing high efficient effort are

$$t_i(Q) = \begin{cases} 
  h \left( \frac{\psi}{p(n)} \right) & \text{if } Q = \overline{Q} \\
  0 & \text{if } 0 \leq Q < \overline{Q}.
\end{cases}$$  \hspace{1cm} (41)

for all $i$.

(ii) The expected team transfer

$$H^T(n, \psi) = p(n)h \left( \frac{\psi}{p(n)} \right)$$  \hspace{1cm} (42)

strictly exceeds the first best transfer $h(\psi)$ for all $n \geq 1$ and $\psi > 0$.

We now consider the relationship between employment $n$ and the optimal team incentive $h(\psi/p)$ and expected transfer $H^T(n, \psi)$.

Proposition 13

(i) The optimal team incentive $h(\psi/p)$ and expected transfer $H^T(n, \psi)$ are increasing in $n$ for all $n \geq 1$ and $\psi > 0$.

(ii) $h(\psi/p) \to \infty$ and $H^T(n, \psi) \to \infty$ as $n \to \infty$ for all $\psi > 0$.

In the previous section the second best incentive and expected transfer were decreasing in employment and converged to their first best levels as the moral hazard problem vanished in the limit. In this section an increase in employment reduces the probability that team output is high which reduces the expected reward for high efficient effort. To implement the latter the principal must therefore increase the incentive $h(\psi/p)$ and the expected transfer $H^T(n, \psi)$ because of the additional risk. This implies a size-wage differential where wages increase with firm size. In the limit the moral hazard problem grows without bound and the optimal team incentive and expected transfer explode.

Proposition 14

(i) The team marginal cost of employment is given by

$$C^T_n(n, \psi) = H^T(n, \psi) \Delta^T(n, \psi),$$  \hspace{1cm} (43)

where $\Delta^T(n, \psi) = 1 + \epsilon_p (1 - \epsilon_h)$ and $\epsilon_p = np^I/p$ is the elasticity of $p(n)$ with respect to $n$. 

20
(ii) $\Delta^T > 1$ for all $n \geq 1$ and $\psi > 0$ and $C^T_n \to \infty$ as $n \to \infty$.

The decomposition $C^T_n = H^T \Delta^T$ is similar to that in the previous section. In both settings each additional hire must be paid the expected transfer $H^I$ or $H^T$ and an increase in employment alters the extent of the moral hazard problem and the corresponding expected transfer as captured by $\Delta^I$ or $\Delta^T$. In the previous section $\Delta^I < 1$ because greater employment reduced the moral hazard problem but in this section it grows so $\Delta^T > 1$ because $1 - \epsilon_h < 0$ and $\epsilon_p < 0$. Since $\Delta^T > 1$ and the expected transfer explodes with employment the marginal cost of employment likewise explodes.

Let $n^T(B, \psi^{-1})$ be the optimal team employment correspondence.

**Proposition 15** If $0 \leq \gamma < 1$ and

$$\left(p^\prime\right)^2 \left(2 + \frac{\psi h''}{p h'}\right) > pp''$$

for all $n \geq 1$ and $\psi > 0$ then $n^T(B, \psi^{-1}) \neq \emptyset$ and every $n \in n^T(B, \psi^{-1})$ is finite.

Intuitively, with probability $p(n)$ team output is high $Q = Bn^{1+\gamma}$ and the wage bill of the principal is $nh(\psi/p)$. Factoring out the $n$, expected team profit depends on the relative magnitudes of team output per head $Bn^\gamma$ and the individual transfer $h(\psi/p)$. One can verify that (44) holds in our Main Example. This condition ensures that $p(n)$ is not too convex in $n$ so the moral hazard problem grows sufficiently rapidly with employment. This implies that $h(\psi/p)$ is increasing and strictly convex so the optimal team size is bounded because $n^\gamma$ is concave for all $0 \leq \gamma < 1$.

Recall the definition of $n^F(B, \psi^{-1})$ immediately prior to Proposition 9.

**Proposition 16**

(i) If $\pi^F(B, \psi^{-1})$ is finite then $n^T(B, \psi^{-1}) \neq \emptyset$ and $n \leq n^F(B, \psi^{-1})$ for all $n \in n^T(B, \psi^{-1})$.

(ii) If the first best employment level $n^F(B, \psi^{-1})$ is unique and finite then $n \leq n^F(B, \psi^{-1})$ for all $n \in n^T(B, \psi^{-1})$.

Under Assumption 4 the optimal team size is limited by moral hazard relative to the first best firm when the first best employment level is unique. Note that this statement encompasses the case where the first best firm is infinite.\footnote{In this section $\epsilon_p$ is the elasticity of $p$ with respect to $n$ whereas in the previous section $\epsilon_\pi$ is the elasticity of $\pi$ with respect to $\rho$. This explains why $\Delta^I = 1 - \epsilon_\pi (1 - \epsilon_h)$ (note the minus sign) whereas $\Delta^T = 1 + \epsilon_p (1 - \epsilon_h)$.} We first discuss (i). Consider an increase in employment

\footnote{I.e., when the first best firm is infinite clearly the optimal team size must be at least weakly smaller.}
above the maximum first best level $\pi^F$ (omitting the relevant arguments). The effect of the increase in employment on expected revenue is the same for the team and first best firm. The cost to the first best firm is $h(\psi)$ for each additional hire above $\pi^F$. But $\Delta^T > 1$ and $H^T > h(\psi)$ imply that

$$C_n^T = H^T \Delta^T > h(\psi) = C_n^F.$$  \hspace{3cm} (45)

Since the first best firm will not increase employment above $\pi^F$ neither will the optimal team. Note that this result ensures that the optimal team size is finite under alternative conditions relative to Proposition 15. The second result is an immediate corollary to the first.

Our final result shows that under the above assumptions the comparative statics of the teams problem are straightforward. Let $\Omega^T$ be the set of all $(B, \psi^{-1}) \in \mathbb{R}^2_{++}$ such that $0 \notin n^T(B, \psi^{-1})$ and define $\rho^T$, $\pi^T$, $e^T$, $l^T$, $\tau^T$, $q^T$, and $Q^T$ as in (16)-(21) except that $n^F$ is replaced by $n^T$.

**Proposition 17** Assume the hypotheses of Proposition 15.

(i) The team employment correspondence $n^T: \Omega^T \rightarrow \mathbb{N}$ is strongly increasing.

(ii) The optimal proportion $\rho^T$ of tasks is strongly decreasing on $\Omega^T$.

(iii) The probability $p^T$ of high team output, production effort $e^T$, investments $l^T$ in human capital, total effort $\tau^T$ per task, and high individual $q^T$ and team $Q^T$ output are all strongly increasing on $\Omega^T$.

We use our Main Example to illustrate the comparison between the optimal team size and the first best firm in Proposition 16 and the comparative statics results in Proposition 17.

**Main Example**

We allow $n$ to be a continuous variable but we impose the condition

$$B \geq \frac{2D^2\psi^2}{\gamma}$$  \hspace{3cm} (46)

to rule out solutions $0 < n < 1$ where $\rho > 1$. Under this condition, (14) and (15) become

$$n^F = \left( \frac{B\gamma}{D\psi^2} \right)^\frac{1}{\gamma^*} \geq 1$$  \hspace{3cm} (47)
when $0 < \gamma < 1$ and $n^F = \infty$ when $\gamma = 1$. Finally, $n^F = \infty$ for all $\gamma > 1$ as before. We now consider optimal team employment. Expected team profit is

$$\Pi^T = pBn^{1+\gamma} - nph(\psi/p) = \frac{B}{D} n^{\gamma} - D\psi^2 n^2.$$  \hfill (48)

The optimal team size is therefore finite

$$n^T = \left( \frac{B\gamma}{2D^2\psi^2} \right)^{\frac{1}{2-\gamma}} \geq 1$$ \hfill (49)

when $0 < \gamma < 2$ and infinite when $\gamma > 2$. Note that the expression in (49) is increasing in $B$ and decreasing in $\psi$ as indicated in Proposition 17. If $\gamma = 2$ then $n^T = \infty$ if the inequality in (46) is strict or indeterminate if it holds as an equality. We now compare (47) and (49) when $0 < \gamma < 1$. Since $D > e \approx 2.718$ we have

$$1 \leq \frac{B\gamma}{2D^2\psi^2} < \frac{B\gamma}{D\psi^2},$$ \hfill (50)

$\frac{1}{1-\gamma} > 1$, and $\frac{1}{2-\gamma} < 1$. It follows that $n^T < n^F$ for all $0 < \gamma < 1$. In summary, $n^T \leq n^F$ for all $\gamma > 0$. If $0 < \gamma < 1$ then $n^T < n^F$ and both are finite. If $1 \leq \gamma < 2$ moral hazard limits the size of the optimal team in the extreme sense that $n^F$ is infinite but $n^T$ is finite. \hfill □

6 Conclusion

In this paper we developed a teams model where team size is endogenous to study the relationship between moral hazard and employment. In some respects the model is similar to the textbook moral hazard model with one agent except that here we make the natural assumption that the probability that individual output is high when the agent chooses high efficient effort declines with the share of tasks performed by the agent. We showed that when individual output is contractible the relationship between employment and moral hazard is ambiguous. Furthermore, incentives and employment are strategic substitutes and incentives and wages decline with employment. While there is some evidence in support of the former, the latter prediction has been consistently refuted by evidence that wages increase with the size of the firm.

When only team output is contractible we provided necessary and sufficient conditions such that moral hazard grows with the size of the firm. In this case incentives and employment are strategic complements and incentives and wages increase with employment. The latter prediction is consistent with evidence on the existence of a size-wage differential. We provided sufficient
conditions such that the optimal team size is finite and smaller than first best. Our Main Example shows that moral hazard can limit the size of the firm in the extreme sense that the optimal team size is finite whereas the first best firm is infinite. In addition to these results we believe our model is tractable enough to address a wide variety of issues in contract theory and the economics of organizations where it is important to model both incentive contracting and the employment decision simultaneously.

7 Appendix

Main Example. We show that the Main Example satisfies Assumptions 1 and the hypotheses of Proposition 11. We first note that \( \pi(\rho) = e^{\rho \log(\rho/D)} \). Using L'Hôpital’s rule,

\[
\rho \log \frac{\rho}{D} = \frac{\log \frac{\rho}{D}}{\frac{1}{\rho}} \to 0 \tag{51}
\]
as \( \rho \to 0 \) so \( \pi \) is continuous at \( \rho = 0 \). It is smooth (i.e., infinitely differentiable) on \((0,1]\) with derivative

\[
\pi' = \pi \left[ 1 + \log \left( \frac{\rho}{D} \right) \right] < 0 \tag{52}
\]
for all \( D > \epsilon \). Note from (52) that \( \pi' \to -\infty \) as \( \rho \to 0 \). Furthermore,

\[
e_{\pi} = \frac{\rho \pi'}{\pi} = \rho + \rho \log \left( \frac{\rho}{D} \right) > \rho \log \left( \frac{\rho}{D} \right) = \log \pi \tag{53}
\]
for all \( \rho \in (0,1] \). ■

Proof of Proposition 1. Let \( T_i = 1 \). Given \( \tau_{is} \geq 0 \) for some task \( s \in S_i \), the optimal allocation of \( \tau_{is} \) between \( e_{is} \) and \( l_{is} \) is

\[
ee_{is} = \frac{\tau_{is}}{1 + \gamma} \quad \text{and} \quad l_{is} = \frac{\gamma \tau_{is}}{1 + \gamma}. \tag{54}
\]
In that case,

\[
q_{is} = \frac{A}{\gamma} \left( \frac{\gamma \tau_{is}}{1 + \gamma} \right)^{1+\gamma} = B \tau_{is}^{1+\gamma}. \tag{55}
\]
Let \( \tau_{is} \) be an integrable function of \( s \in S_i \) which satisfies \( \tau_{is} \geq 0 \) and (3) with \( T_i = 1 \). Let \( I_i = \inf_{s \in S_i} \tau_{is} \). Suppose there exists a subset \( S_i' \) of \( S_i \) of positive measure such that \( \tau_{is} > I_i \) for all \( s \in S_i' \). In that case the principal could adjust \( \tau_{is} \) such that \( \tau_{is} = I_i \) for all \( s \in S_i' \) and then increase \( I_i \) slightly so as to preserve (3). From (2) it follows that \( \tau_{is} \) must be almost everywhere constant at the optimum. It is therefore optimal to divide \( T_i = 1 \) evenly \( \tau_{is} = 1/\rho \) across tasks. ■
Proof of Proposition 2. The principal’s problem is to choose \( t_i \geq 0 \) and \( t_i \geq 0 \) to minimize

\[
\pi(\rho)t_i + [1 - \pi(\rho)]t_i
\]

subject to the participation constraint (11). Since \( u \) is not necessarily differentiable at \( t_i = 0 \) we first consider the case where the domain of the objective function and participation constraint is \( (t_i, t_i) \in \mathbb{R}_+^2 \). Note that the objective function is linear and the constraint is strictly concave so the standard concave programming theorem applies. The Lagrangean and first-order conditions are

\[
\mathcal{L} = -\pi \bar{t} - (1 - \pi)\bar{t} + \lambda \left[ \pi u(\bar{t}) + (1 - \pi)u(t) - \psi \right]
\]

\[
\mathcal{L}_t = -\pi + \lambda \pi u'(\bar{t}) = 0
\]

\[
\mathcal{L}_\bar{t} = -(1 - \pi) + \lambda (1 - \pi)u'(t) = 0
\]

\[
\mathcal{L}_\lambda = \pi u(\bar{t}) + (1 - \pi)u(t) - \psi \geq 0,
\]

where we omit the argument \( \rho \) in \( \pi \), drop the \( i \) subscript, and omit the complementary slackness condition. From the first two first-order conditions we observe that \( \lambda > 0 \) and \( \bar{t} = \bar{t} \) because \( u'' < 0 \). Since the participation constraint binds we have that \( \bar{t} = \bar{t} = h(\psi) \). We now check the boundary. One potential solution is \( t = 0 \) and \( t = h(\psi/\pi) \). We show that this entails higher expected cost in the proof of Proposition 6. Another potential solution is \( \bar{t} = 0 \) and \( \bar{t} = h(\psi/\pi) \) but

\[
\pi(0) + (1 - \pi)h\left(\frac{\psi}{1 - \pi}\right) = \pi h(0) + (1 - \pi)h\left(\frac{\psi}{1 - \pi}\right) > h\left[\pi(0) + (1 - \pi)\frac{\psi}{1 - \pi}\right] = h(\psi)
\]

since \( h \) is strictly convex. ■

Proof of Proposition 3. Write expected profit as \( \Pi^F = n \left[ p(n)Bn^\gamma - h(\psi) \right] \). The statement in (ii) is clear. If \( p(n)n^\gamma \rightarrow 0 \) as \( n \rightarrow \infty \) then there exists an \( N \in \mathbb{N} \) such that \( \Pi^F < 0 \) for all \( n \geq N \). The principal therefore chooses \( n \) from a finite bounded subset of \( \mathbb{N} \) which proves (i). ■

Proof of Proposition 4. If \( p(n)n^\gamma \rightarrow 0 \) as \( n \rightarrow \infty \) then for \( (B, \psi^{-1}) \in \Omega^F \) the first best firm is finite but non-zero. The proof is an application of Theorem 2.3 in Vives (1999, p. 26) which we state below for convenience.

**Theorem 2** Let \( X \) be a lattice, \( T \) a partially ordered set, and \( S \) a nonempty sublattice of \( X \). If \( f(x,t) \) is supermodular in \( x \) on \( X \) for all \( t \in T \) with strictly increasing differences on \( X \times T \) then
arg\ max_{x \in S} f(x, t) \text{ is strongly increasing in } t.

Let

\[ g(n) = \begin{cases} 
  p(n)n^{1+\gamma} & n \in \mathbb{N} \\
  0 & n = 0 
\end{cases} \tag{62} \]

and

\[ N_g = \left\{ n \in \mathbb{N} \mid g(n) > g(n-1) \right\} \tag{63} \]

The principal never chooses \( n \) in the complement of \( N_g \) in \( \mathbb{N} \) because otherwise she could reduce \( n \) and her costs would decrease while her expected revenue would weakly increase. Note that \( N_g \) is nonempty because the first best firm is finite but non-zero. In Theorem 2 we set \( X = S = N_g \) and \( T = \Omega^F \) for the parameters \((B, \psi^{-1})\). Since every function of one variable on a lattice is supermodular we need only prove strictly increasing differences. Let \( B \geq B' \) and \( \psi^{-1} \geq (\psi')^{-1} \) with at least one inequality strict. The latter implies that \( \psi' \geq \psi \) and \( h(\psi') \geq h(\psi) \). The result follows because

\[ \Pi^F(n \mid B, \psi) - \Pi^F(n \mid B', \psi') = (B - B')p(n)n^{1+\gamma} + [h(\psi') - h(\psi)]n \tag{64} \]

is increasing in \( n \) on \( N_g \).

**Proof of Proposition 5.** The principal’s problem is

\[ \min_{t_i(q_i)} \pi(\rho)t_i(\bar{q}) + [1 - \pi(\rho)]t_i(0) \tag{65} \]

subject to \( t_i(q_i) \geq 0 \) for all \( 0 \leq q_i \leq \bar{q} \), (23), (26), and (27). We first consider the problem neglecting the constraints (27). Since \( u(t_i) \geq 0 \) for all \( t_i \geq 0 \), (26) implies (23). We can therefore drop the latter constraint because it is redundant. Since (26) reduces to

\[ u(t_i(\bar{q})) - u(t_i(0)) \geq \frac{\psi}{\pi(\rho)} \tag{66} \]

the solution is clear. The omitted constraints are then satisfied by \( t_i(\hat{q}_i) = 0 \) for all \( 0 \leq \hat{q}_i < \bar{q} \).

**Proof of Proposition 6.** Since the participation constraint (23) binds and \( u \) is strictly concave,

\[ \psi = \pi(\rho)u(t_i(\bar{q})) + [1 - \pi(\rho)]u(t_i(0)) < u[\pi(\rho)t_i(\bar{q}) + [1 - \pi(\rho)]t_i(0)] \tag{67} \]
Since $h$ is increasing,

$$h(\psi) < \pi(\rho)t(\pi) + \left[1 - \pi(\rho)\right]t(0) = \pi(\rho)h \left[\frac{\psi}{\pi(\rho)}\right] = H^I(\rho, \psi). \quad (68)$$

Assumptions 2 imply $h$ is twice continuously differentiable on $\mathbb{R}_{++}$ with $h' > 0$ and $h'' > 0$. Since $h$ is strictly convex,

$$h'(u_1)(u_2 - u_1) < h(u_2) - h(u_1) \quad (69)$$

for all $u_1 \neq u_2$. Setting $u_1 = \psi/\pi$ and $u_2 = 0$,

$$h(\psi/\pi) - \frac{\psi}{\pi} h'(\psi/\pi) < 0 \quad \text{and} \quad 1 - \frac{\psi}{\pi} \frac{h'(\psi/\pi)}{h(\psi/\pi)} < 0. \quad (70)$$

Differentiating $H^I(\rho, \psi)$ with respect to $\rho$,

$$H^I_\rho = H^I_x \pi' = \left(h - \frac{\psi}{\pi} h'\right) \pi' > 0. \quad (71)$$

**Proof of Proposition 7.** Differentiating $C^I$ with respect to $n$,

$$C^I_n = nH^I_\rho \left(-\frac{1}{n^2}\right) + H^I = H^I - \frac{1}{n} H^I_\rho = \pi h - \frac{1}{n} \left(h - \frac{\psi}{\pi} h'\right) \pi' \quad (72)$$

$$= \pi h \left[1 - \frac{1}{n} \pi' \left(1 - \frac{\psi h'}{\pi h}\right)\right] = H^I \Delta^I. \quad (73)$$

The result in (ii) follows because $\epsilon_\pi \to 0$ and $1 - \epsilon_\pi$ is bounded by Assumptions 2. ■

**Proof of Proposition 8.** Write expected profit as $\Pi^I = n \left[p(n)Bn^\gamma - H^I(n, \psi)\right]$. Since $H^I(n, \psi) \to h(\psi)$ as $n \to \infty$ for all $\psi > 0$ the result follows. ■

**Proof of Proposition 9.** If the first and second best firms are infinite the result holds. Assume both are finite. Let $\bar{n}^I_+ = \max_n n^I_+$ and $\bar{n}^F = \max_n n^F$, where we omit the arguments $B$, $\psi$, and $\psi^{-1}$. Note that we could have $\bar{n}^I_+ = \infty$. If $\bar{n}^I_+ \leq \bar{n}^F$ the conclusion follows. Assume $0 \leq \bar{n}^F < \bar{n}^I_+$ and suppose the second best firm chooses $0 \leq \bar{n}^F < n \leq \bar{n}^I_+$, where $n = \bar{n}^F + k$ for some $k \in \mathbb{N}$. This increase in employment increases $C^F$ by $kh(\psi)$ and reduces $\Pi^F$. Since the effect on expected revenue is the same for both institutions and

$$C^I(n, \psi) - C^I(\bar{n}^F, \psi) = \int_{\bar{n}^F}^{n} C^I_n(n, \psi) dn \geq \int_{\bar{n}^F}^{n} h(\psi) dn = kh(\psi) \quad (74)$$
the increase in employment also reduces $\Pi^I$. ■

**Proof of Proposition 10.** Since $p(n)n^\gamma \to 0$ as $n \to \infty$ second best employment levels exist and are finite. Note that each $n \in n^I(B,\psi^{-1})$ satisfies $n \in \mathbb{N}_g$ [see (63)] because $C^I_n > 0$ for all $n \geq 1$ and $\psi > 0$. Let $B \geq B'$ and $\psi^{-1} \geq (\psi')^{-1}$, with at least one inequality strict. To establish strictly increasing differences, we must show that

$$\Pi^I(n|B,\psi) - \Pi^I(n|B',\psi') = (B - B')p(n)n^{1+\gamma} + n\pi(\rho)\left\{h\left[\frac{\psi'}{\pi(\rho)}\right] - h\left[\frac{\psi}{\pi(\rho)}\right]\right\}$$

(75) is strictly increasing on $\mathbb{N}_g$. The first term satisfies this by the definitions of $g$ [see (62)] and $\mathbb{N}_g$. The second term is strictly increasing in $n$ iff $C^I_{n^\psi} > 0$ as assumed. ■

**Proof of Proposition 11.** We first note that $p(n)$ is decreasing in $n$ iff $\log p(n) = (1/\rho)\log \pi(\rho)$ is increasing in $\rho$. Differentiating with respect to $\rho$,

$$\frac{1}{\rho} \frac{\pi'}{\pi} - \frac{1}{\rho^2} \log \pi(\rho) > 0 \iff \frac{\rho \pi'}{\pi} > \log \pi(\rho)$$

(76) which proves (i). Next we note that $p(n) \to 0$ as $n \to \infty$ iff

$$\log p(n) = n \log \pi(1/n) = \frac{\log \pi(1/n)}{1/n} \to -\infty.$$  

(77)

Using L'Hôpital's rule,

$$\lim_{n \to \infty} \frac{\log \pi(1/n)}{1/n} = \lim_{n \to \infty} \frac{\pi'(1/n)}{\pi(1/n)} = \lim_{n \to \infty} \pi'(1/n)$$

(78) which proves (ii). ■

**Proof of Proposition 13.** Statement (i) follows from the derivative

$$\frac{dH^T(n,\psi)}{dn} = p'\left(h - \frac{\psi}{p} h'\right) > 0.$$  

(79)

We now prove (ii). Since $p(n) \to 0$ and $h$ is increasing and strictly convex, $h \to \infty$ as $n \to \infty$ and

$$\lim_{n \to \infty} ph(\psi/p) = \lim_{n \to \infty} \frac{h(\psi/p)}{\frac{1}{p}} = \lim_{n \to \infty} \psi h'(\psi/p) = \infty$$

(80) by L'Hôpital's rule and Assumptions 2. ■
Proof of Proposition 14. Differentiating $C^T$ with respect to $n$,

$$
C^T_n = nH^T_n + H^T = np' \left( h - \frac{\psi}{p} h' \right) + ph
$$

$$
= ph \left[ 1 + \frac{np'}{p} \left( 1 - \frac{\psi}{p} h' \right) \right] = H^T \left[ 1 + \epsilon_p (1 - \epsilon_h) \right]
$$

which proves (43). Since $1 - \epsilon_h < 0$ and $p' < 0$ and $\epsilon_p < 0$ for all $n \geq 1$ we have $\Delta^T > 1$ for all $n \geq 1$. Since $H^T \to \infty$ as $n \to \infty$ this proves (ii). ■

Proof of Proposition 15. The condition (44) implies that $h(\psi/p)$ is strictly convex in $n$ because

$$
\frac{d h(\psi/p)}{dn} = -\frac{\psi}{p^2} p' h'
$$

$$
\frac{d^2 h(\psi/p)}{dn^2} = \frac{\psi}{p^2} h' \left[ \left( p' \right)^2 \left( 2 + \frac{\psi}{p} h'' \right) - p'' \right].
$$

Write expected team profit as

$$
\Pi^T(n \mid B, \psi) = p(n)Bn^{1+\gamma} - np(n)h \left[ \frac{\psi}{p(n)} \right] = p(n)Bn^{1+\gamma} \left\{ 1 - \frac{h \left[ \frac{\psi}{p(n)} \right]}{Bn^{1-\gamma}} \right\}.
$$

We focus on the term $h/n^\gamma$. Since the numerator and denominator $\to \infty$ as $n \to \infty$ we can apply L'Hôpital’s rule to conclude that

$$
\lim_{n \to \infty} \frac{h(\psi/p)}{n^\gamma} = \lim_{n \to \infty} \frac{1}{\gamma} \frac{d h(\psi/p)}{dn} n^{1-\gamma} = \infty
$$

because the derivative is positive and increasing in $n$ and $0 \leq \gamma < 1$ implies $n^{1-\gamma} \to \infty$. It follows that there exists an $N \in \mathbb{N}$ such that expected team profit is negative for all $n \geq N$. ■

Proof of Proposition 16. For all $n \geq \pi^F(B, \psi^{-1}) > N_{p'}$ we have $p'(n) < 0$, $\Delta^T > 1$, and $H^T > h(\psi)$. It follows that $C^T_n = H^T \Delta^T > C^F_n$ for all $n \geq \pi^F(B, \psi^{-1})$. The rest of the proof is similar to that for Proposition 9. ■

Proof of Proposition 17. From Proposition 15, $n^T(B, \psi^{-1}) \neq \emptyset$ and every $n \in n^T(B, \psi^{-1})$ is finite for all $(B, \psi^{-1}) \in \Omega^T$. Since $C^T_n > 0$, $n^T(B, \psi^{-1}) \subseteq \mathbb{N}_g$ for all $(B, \psi^{-1}) \in \Omega^T$ (see the proof of Proposition 4) and we must prove increasing differences on the latter set. Let $B \geq B'$ and $\psi' \geq \psi$. 

29
with at least one inequality strict. We need to show that

$$\Pi^T(n \mid B, \psi) - \Pi^T(n \mid B', \psi') = (B - B')pn^{1+\gamma} + np\left[h(\psi'/p) - h(\psi/p)\right]$$

(87)

is increasing in \(n\) on \(\mathbb{N}_g\). The term \(pn^{1+\gamma}\) is increasing in \(n\) on \(\mathbb{N}_g\). The second term is increasing in \(n\) iff \(C^T_{n\psi} > 0\) iff \(C^T_{\psi} = nh'(\psi/p)\) is increasing in \(n\) which holds because \(p\) is decreasing for all \(n \geq 1\). ■

8 References


