Incentives for Spot Market Labor When Output is Unverifiable*

J. Aislinn Bohren
University of Pennsylvania

Troy Kravitz
University of California, San Diego

February 15, 2013

Abstract

This paper studies a novel contracting environment plagued by moral hazard. The setting differs from traditional agency theory along two dimensions: no signal about whether effort was exerted is available, and the firm is unable to threaten workers with negative wage payments. The paper characterizes the firm's optimal incentive organization. The firm hires multiple workers to perform some tasks as a monitoring device. Consequently, monitoring is determined endogenously. The optimal mechanism features two distinctive features absent in standard contracting solutions. First, the firm monitors with nontrivial probability by hiring additional workers to complete a subset of the tasks it assigns to any agent. Second, the firm bundles multiple tasks together to reduce the per-task cost of monitoring. By assigning more tasks to each worker and conditioning wage payment for any task on satisfactory performance on all tasks, the firm approximates its first-best payoffs even when firing and large punishments are unavailable.

*We thank Nageeb Ali, Cécile Aubert, Gordon Dahl, Roger Gordon, Kevin Lingerfelt, Mark Machina, David Miller, David Rahman, Joel Sobel, Ross Starr and Joel Watson.

1 Introduction

Technological advancement is changing the way modern companies interact with their workforce. Traditional in-house employees only recently gave way to outsourced workers contractually tied to the firm. Further changes are afoot: new information technology creates the opportunity for firms to access a flexible and inexpensive pool of workers on-demand. One example is crowdsourcing, where firms hire temporary workers through open calls online. Millions of potential employees are available around the clock and able to start work immediately. Neither a preexisting relationship nor a continuing relationship is required. In the absence of conventional methods of supervising employees, tapping into this global workforce presents a host of new incentive issues.

Much of the allure of spot labor markets is that interaction between the firm and its employees is minimal. This is also a liability as reputation mechanisms have little bite. Workers are compensated for their effort, but the exertion of effort is unobservable and no direct signal about whether effort was exerted is available. The firm must guard against shirking, but how?

Traditional agency theory offers limited guidance for firms interacting in these spot markets. Economic theories of crime, like those of Becker (1968) and Kolm (1973), for example, suggest optimal enforcement regimes combining infinitesimal monitoring with arbitrarily harsh punishments. Mirrlees (1974) shows that this principle applies broadly in principal-agent settings. Yet actual firms operate quite differently from the theoretical recommendations (Dickens, Katz, Lang, and Summers, 1989). Considerable resources are devoted to monitoring employees and firms are unwilling or unable to impose sufficiently harsh punishments to stamp out undesirable employee behavior. This suggests the problem is appropriately framed with constraints on the firm's available punishments.

We study the challenges faced by a firm hiring in a spot market by viewing the situation within a principal-agent context. The firm has a series of items to categorize. Each item can be thought of as having a state that is unknown to the firm but that can be observed by a worker who exerts costly effort. Worker payoffs are independent of the state and protected by limited liability clauses, effort is non-contractible, and output is unverifiable. Each worker is offered a contract consisting of tasks to be completed, a monitoring protocol specifying how many workers are to be employed for each task, a verification technology to review output, and a payment schedule.

The setting is similar to multilateral principal-agent problems except that the firm lacks a direct signal about worker quality. In the agency literature, the principal typically receives a direct signal about output that is related to unobserved effort. With no such signal available, the firm draws inferences about the quality of worker input by hiring multiple workers to complete the same task. In effect, the firm generates a costly signal about effort by monitoring workers with other workers. Correlation in agents' information is used to discipline behavior without resort to large negative payments as threats.

The main result of the paper derives the firm's optimal organization of incentives. The optimal mechanism exhibits two features absent in standard contracting solutions: endogenous monitoring and bundling. Effort cannot be induced without hiring additional workers. The firm assigns an agent a collection of tasks to complete. It then hires another worker to duplicate a subset of these tasks and pays the agent only for satisfactory performance on all tasks. Producing evidence of shirking is costly for the firm and the amount and efficacy of employee monitoring is determined endogenously.

Wages depend on the frequency of monitoring: higher wages are needed as the firm monitors less frequently. At the optimum, the firm employs monitors frequently enough to keep wages down but infrequently enough to prevent excessive duplication of assignments. The mechanism is wasteful compared to the observable effort benchmark: not only must additional workers be hired, but each must be paid a wage greater than his cost of effort.

Alchian and Demsetz (1972) suggest that there should be specialization in monitoring. This is not the case in our setting: the optimal incentive organization of the firm treats workers symmetrically. The distinction between monitors and subordinates is solely expository. Armen monitors Harold while, at the same time, Harold monitors Armen. No hierarchy of monitors is created.

Information aggregation is often emphasized as a motivation for hiring multiple agents (Aghion, Bolton, Harris, and Jullien, 1991; Grossman, Kihlstrom, and Mirman, 1977; Rothschild, 1974). In the canonical setting of statistical decisionmaking, a gambler tests a slot machine multiple times to acquire improved information about the machine's unknown odds. The question is when to move on to another machine if the gambler's objective is to maximize winnings. As both the principal and the agent, the gambler does not need to provide incentives for the agent to undertake the principal's desired behavior: experimentation is solely for learning reasons.

In the model of Section 4, by contrast, there is no learning justification for the duplication of tasks: multiple workers are employed solely for incentive reasons. The setting without noise can be interpreted as the opposite pole to the setting of Robbins (1952). The paper thus provides a complementary explanation for why a principal may consult multiple agents before taking a decision.

Another contribution of this paper is to highlight the role played by combining multiple tasks into one job. Bundling tasks together permits the principal to reduce the cost of monitoring. Specifically, the firm assigns multiple tasks to each worker and hires another worker to perform only a few of these tasks. The worker's payment is then conditioned on his performance on the tasks that are monitored.

The firm cannot punish workers by assessing penalties for poor performance. The most it can do is withhold wages. So while the firm generates incentives by stochastically duplicating an agent's work, tying the agent's wage for one task to his performance on all tasks strengthens these incentives. Monitoring and bundling are strategic substitutes: the firm hires monitors less frequently as the number of tasks assigned to an agent increases. The efficiency loss relative to the contractible effort benchmark vanishes asymptotically.

The theoretical setting is presented within the context of spot labor markets in general and the online crowdsourcing marketplace in particular. Spot markets for labor are already big business: in crowdsourcing, where firms and workers interact through the web, worker earnings are in the billions, the revenues of vendors matching firms to workers were estimated at $500 million in 2009 (Frei, 2009) and venture capital firms injected almost $300 million in 2011 (Sanders, 2011). It is not surprising that industry insiders believe one-third of the global workforce could be hired online by 2020 (Vanham, 2012).

The mechanisms we identify offer potentially significant improvement over those currently in practice. Firm can reduce their monitoring expenditures by structuring contracts so that individual workers check each other. By shifting from piece-rate payment schedules to schemes requiring satisfactory performance on all tasks, firms can recreate the same incentives at lower cost. The theoretical performance improvements can be empirically tested through field experiments carried out online.

The main findings of the paper apply to other settings as well. For example, considering a multidimensional chore instead of a series of tasks, the firm monitors individual components of the chore and punishes workers across all dimensions for poor performance on any dimension. Tax auditing practices, hiring rules-of-thumb and the monitoring of rider provisions are examples of the main ideas of the paper applied to this setting.

The paper proceeds as follows. Related literature is discussed in Section 2 and a representative example is presented in Section 3. The example isolates key properties of the setting. The formal model is introduced in Section 4 and the firm's optimal incentive organization is derived in Section 5. Discussion and extensions follow in Section 6. All proofs are relegated to Appendix B following the conclusion. Appendix A contains a brief overview of a setting of interest, the crowdsourcing marketplace.

2 Literature

In the standard agency model, an agent selects action a and the principal observes s, an informative signal about a. The principal cares about output y, which depends on a and s. The principal's problem is to generate incentives for a particular a^* using only a function of s.[1] Actions can be interpreted broadly; a^* is commonly an optimal level of investment or effort. The signal s may be something like sales volume or peer evaluations. As long as the conditional distribution of the signal $p(s|a)$ varies with a, the firm is able to align incentives so that a^* is the employee's voluntary action choice.[2]

The characterization of optimal contracts in the general setting is limited.[3] Instead, the literature provides a host of elegant solutions for specific settings. Many of the mechanisms identified are variations on a common dynamic: if for each a there exists s such that $p(s|a^*) \neq p(s|a)$, the firm is able to statistically discriminate among worker actions. When s is realized, the firm inserts a wedge until a is no longer attractive. Bad signals – those that are more frequent when a is taken instead of a^* – are punished while good signals are rewarded.

Sometimes the observable signal s only reveals aggregate information about actions instead of individual action choices. The firm is now in a multilateral contracting environment and must guard against the free-rider problem. Holmstrom (1982) emphasizes the role of group penalties: all workers are punished whenever bad signals obtain. Group penalties are natural here since the firm cannot discern which worker deviated. They continue to play a role elsewhere, especially when identifying the deviator is more costly than identifying that a deviation occurred.

A practical, though not theoretical, complication arises if the conditional distribution

[1]Mirrlees (1975), Mirrlees (1976) and Holmstrom (1979) are seminal papers in the literature on moral hazard.

[2]This often introduces a distortion away from the optimal contract with contractible effort since the worker needs to be compensated for the risk in his wage payment.

[3]See Bolton and Dewatripont (2005, Ch. 4).

of s given a is very similar to that given a^*. Here is it difficult for the firm to distinguish whether a or a^* was chosen by the worker. In order to dissuade a, the firm must threaten large punishment in the off-chance it believes a more likely than a^*. Larger and larger punishments are required as the signals triggering punishment become more rare.

The incentive contracts above break down if the firm is limited in how severely it can punish an employee. One response is that instead of punishing a worker by paying him $-x < 0$ when a bad signal obtains, the firm transfers x to the worker at the outset and simply takes away this transfer upon observing a bad signal. Such an arrangement preserves incentives, so if the contract without limited liability is able to induce a^*, then so is the modified contract satisfying limited liability. But note that the firm's expected payments have now increased by x: hiring workers may no longer be attractive at all.

In standard agency models, the principal cannot achieve the first-best equilibrium obtained when actions are contractible. Correlation in agents' valuations or information provides a way for the principal to recover first-best.[4] Viewed in isolation, an agent's action or signal thereof is uninformative about the underlying uncertainty. But viewed in conjunction with those of other agents, the principal is able to draw inferences about the hidden object.

Legros and Matthews (1993) study a partnership problem between the poles of individual and aggregate signals. Each partner privately devotes effort to a common project. Allocating credit for success is impossible except when only one partner is contributing to the project. Randomly with small probability all partners but one shirk, perfectly revealing the action of the working partner, who is then punished according to his effort provision.

The signal is commonly produced without cost as a byproduct of action choices.[5] In this respect, Legros and Matthews (1993) is an example of costly monitoring. Rahman (2012) also considers a costly monitoring setting; in an ideal arrangement, a worker exerts effort and the firm never monitors the worker. This would be the action profile with contractible effort. When payments cannot be made contingent on actions – perhaps because actions are unobservable – this profile cannot be achieved. Rahman (2012) shows that the profile can be approximated arbitrarily well by using the dynamic described earlier. The signal s is now the report of a monitor periodically hired to verify the worker's action.[6] The worker is punished severely whenever the monitor reports that he shirked.

The signal structure can be viewed as the firm's monitoring technology. In much of the literature, regardless of whether signals are a costless byproduct of actions, the monitoring technology is exogenous. Early studies in which monitoring is a choice variable include Becker (1968), Kolm (1973) and Mirrlees (1974);[7] these papers suggest combining infinitesimal monitoring with arbitrarily harsh punishments.

The observed incentive structures of firms bear little resemblance to these arrangements (Dickens, Katz, Lang, and Summers, 1989). Even if arbitrarily harsh punishments are possible in practice, they may still be undesirable since they flatten the penalty gradient between

[4]The surplus extraction literature illustrates, applying the dynamic described earlier to the bid profile from an auction where bidders have correlated valuations of the item being sold (Bose and Zhao, 2007; Cremer and McLean, 1988; McAfee and Reny, 1992; Riordan and Sappington, 1988).

[5]See, for example, the subjective evaluation literature (MacLeod, 2003; Prendergast, 1999).

[6]The relationship of the present paper to Rahman (2012) is discussed further in Section 6.

[7]The efficiency-wage theory of Shapiro and Stiglitz (1984) is sometimes portrayed as an example of endogenous monitoring. See, for example, Bolton and Dewatripont (2005, § 4.1.3).

minor and major offenses (Stigler, 1970). Becker and Stigler (1974), Carr-Hill and Stern (1979) and Carmichael (1985) provide other reasons to question the appropriateness of unlimited liability.

This paper identifies the optimal incentive organization for a firm (i) bound by limited liability and in the absence of repeated interactions and (ii) in a setting where costly monitoring is necessary to produce informative signals. Punishments entail taking away rents when a deviation from a^* is observed. The punishments are strengthened by tying the payment for one part of the job to performance on all parts of the job. This dynamic is similar to that identified by Fuchs (2007) in a repeated setting; there, Fuchs (2007) shows it is optimal for the firm to withhold payment until the final period.[8]

3 Example

Each day users upload thousands of images to an auction website. The firm controlling the website needs to ensure the images are not obscene. A workforce of thousands is available around the clock in an online labor market.

As new item listings are created, the uploaded images pass through a central database for verification. Multiple images can be collected together and sent to workers as a package, but the firm is unwilling to delay approval of prospective listings and so limits each worker to reviewing no more than 10 images at a time.

Workers verify content by viewing the image. This process takes time, the cost of which is \$0.02, but accurately reveals objectionable content without fail. Workers and the firm know that most images are acceptable and that, on average, only 5% will be flagged as obscene. They also know that it is not cost-effective for the firm to verify a worker's suggestion about the status of an image; that's why the firm is hiring them to act on its behalf in the first place.

Payoffs are as follows: the risk-neutral firm earns a payoff of 1 for correctly classifying an image (i.e., forbidding obscene content or approving acceptable content) and a payoff of 0 otherwise, less any wages promised to workers. Risk-neutral workers do not care whether the firm treats an image correctly: payoffs are their received wages less the cost of their effort.

We now consider various settings and identify how the firm optimally structures incentives within each setting. A labor contract consists of the number of images a worker must review, a wage for the job and the conditions for payment.

Observable effort

First, suppose that the firm can observe whether a worker viewed the image. The firm pays a worker their time cost of effort \$0.02 for each image he views and nothing for the images he skips. The firm does not need to hire additional workers for any task and there is no benefit through reduced wages per task to assigning a bundle of images to the worker instead of a solitary image.

[8] Abreu, Milgrom, and Pearce (1991) emphasize the reusability of punishments: one punishment can simultaneously provide incentives across many periods.

Unobservable effort and unlimited liability

Now suppose effort is unobservable. Since the firm cannot verify the accuracy of their recommendations, workers have incentive to bypass viewing the images and randomly report their status. The firm must find a way to monitor the workers it hires.

If the firm assigns the same task to two workers, it can compare their recommendations. Negative wages are permissible with unlimited liability, and so the firm offers the following contract. For each image the worker is assigned, with probability $q > 0$ the firm hires a second agent to report the image's status. The worker is paid \$0.02 unless his recommendation disagrees with that of the second agent, in which case the worker is assessed a penalty of $\$0.02(1 - 1/.05q)$. The firm hires $1 + q$ workers in expectation for each image. As $q \to 0$, the firm hires approximately one worker and compensates him at his cost of effort. Again, there is no benefit to assigning multiple images to a worker.

Unobservable effort and limited liability

The firm can no longer threaten workers with negative wage payments. Suppose the firm follows the incentive structure from above, hiring a second worker with probability q and paying him \$0.02 unless his recommendation differs from that of the second worker. Since the firm cannot penalize the worker with a negative wage, the worker will guess the status of the image and hope that either he is unmonitored or that his guess is correct. Thus, the worker requires a premium above his cost of effort in order to be induced to exert effort. Additionally, the firm must hire additional workers for each task with nontrivial probability. In the current example, a wage of at least \$0.40 per image is required by each worker. These efficiency losses are a general feature of limited liability when effort is unobservable.

The firm can do better by leveraging its ability to assign multiple images to each worker. Two situations are worth highlighting. First, if the firm can make the worker bear the risk that a second agent will not be called upon to verify his recommendation, the firm can offer the worker a contract that pays him if and only if a second agent is hired and verifies his recommendation on every assigned task. The firm promises to hire a second agent independently for each task with probability $q > 0$ and pays the worker $\$0.2/q^{10}(1 - .95^{10})$ if the criteria for payment are satisfied.

As $q \to 0$, the firm hires a second worker for each task less often and fewer workers meet the criteria for payment. The firm's expected payments per image approach $\$0.02/(1 - .95^{10}) \approx 0.05$. When workers are protected by limited liability, the firm can do no better than this arrangement. Note that with unlimited liability there is no benefit to the firm of having workers bear the risk of monitoring.

This incentive structure is implausible. When the firm does not verify the worker's recommendation, then no wages are paid, even though the failure to satisfy the payment criteria is not due to negligence of the worker. The firm has incentive to induce a breach of contract by not monitoring in this setting.

Unobservable effort, limited liability and limited risk

If, instead, workers cannot be made to bear the risk that the firm does not monitor their output, the optimal frequency with which the firm hires a second worker is endogenously determined. At the optimum, the firm's cost per image is just below \$0.10.

In sum, when effort is observable, the firm's cost per image is \$0.02, no additional workers are needed, and there is no benefit to bundling multiple images together. This expected cost can be approximated arbitrarily well with unobservable effort by threatening workers with large fines if their recommendation differs from that of another worker. Additional workers are required with arbitrarily small probability and assigning multiple images to workers provides no benefit to the firm. If the firm continues to hire workers for single images, the firm's expected cost per image increases to \$0.80 when workers are protected by limited liability.

Bundling provides a way for the firm to mitigate the costly effect of workers' limited liability protection. By assigning multiple images to each worker, the firm is able to tie a worker's compensation on one task to successful completion of all tasks. This strengthens the incentives it can offer regardless of whether workers can be made to bear the monitoring risk.

4 Model

4.1 Statement of the problem

A firm is faced with an infinite stream of tasks, and seeks to learn some unknown attribute of each task by delegating this work to agents. For example, a social media firm must moderate the content of user-generated images; the firm hires workers to report whether each image meets content guidelines. An agent can either exert costly effort on an assigned task, in which case the agent learns the true attribute of the task (i.e. whether the image is acceptable), or not exert effort, in which case the agent learns nothing. The agent then sends a report to the firm about the relevant attribute.

Effort is unobservable, an agent's report (in isolation) conveys no information about his effort choice and the agent's preferences are independent of the task's unknown attribute. Therefore, there is no incentive structure that will compel an agent to exert effort when the agent acts in isolation. However, delegating a task to multiple agents and comparing their reports will generate a signal of their effort choices, which can be used to effectively structure incentives.

The firm's problem is to design a labor contract that optimally allocates tasks to workers and compensates workers based on their reports. The firm must decide how many tasks to assign to each worker, how many workers to hire for each task, and a compensation schedule that conditions payment on the profile of reports for each task assigned to a worker. This general multilateral contracting problem is presented formally in the next subsection.

4.2 The formal model

A firm is faced with a countably infinite stream of independent and identical tasks $j = 1, 2, ...$ and has access to a countably infinite pool of workers $i = 1, 2, ...$ The firm can design labor contracts to delegate tasks to workers. There are two formal components to

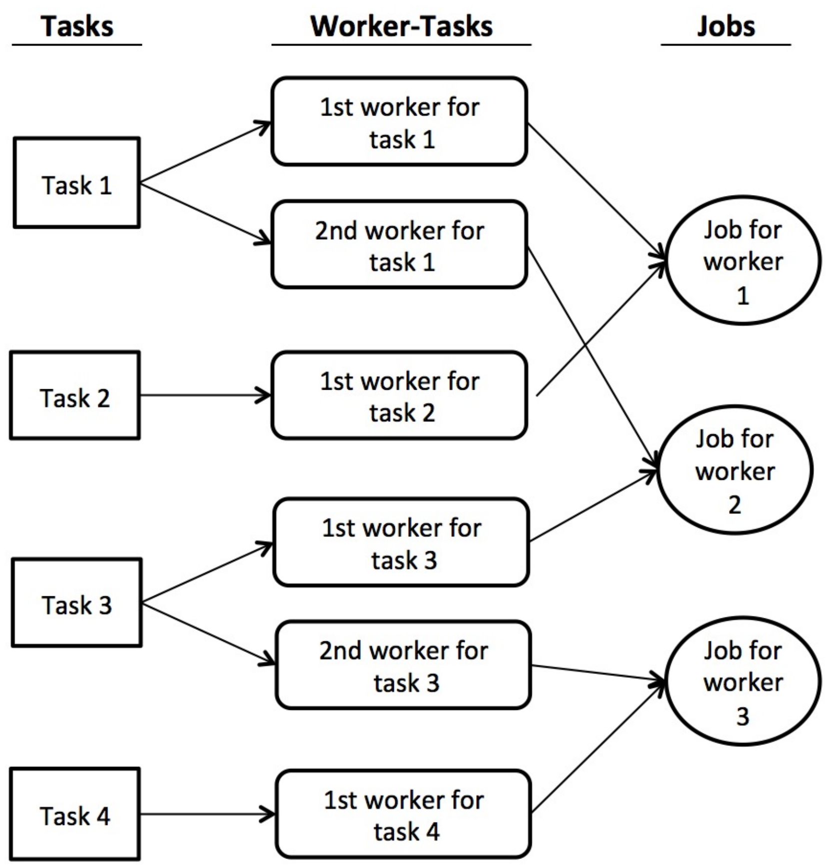

Figure 1: The relationship between tasks, worker-tasks and jobs.

the model: the **task** j and the **job** i, or the set of tasks assigned to worker i.[9] A task can be assigned to multiple workers, and multiple tasks can be assigned to a single worker. We use the phrase **worker-task** to refer to an individual worker's decision problem on a each task within a job (equivalently, the individual decision problem for each worker assigned to a task). Worker-tasks partition both tasks and jobs. Figure 1 describes these components visually.

The Task

On each task j, there are two possible states of the world, $\omega_j \in \Omega \equiv \{0, 1\}$ with common prior belief: [10][11]

$$\pi = Pr(\omega_j = 1) \in [1/2, 1).$$

The firm seeks to match the realized state of the world by choosing an action $A_j \in \Omega$. It receives a payoff of 1 if its action matches the realized state, and a payoff of 0 otherwise.[12] Workers' payoffs are independent of the state.

[9]The following convention is maintained: objects pertaining to tasks are subscripted and objects pertaining to workers are superscripted.

[10]For example, whether an image meets content guidelines.

[11]Section 6.5 extends the model to settings where all tasks are not identical.

[12]Section 6.5 allows firm to have asymmetric payoffs for matching each state

Before selecting an action, the firm can hire workers to learn about the unknown state. For each task j, the firm chooses a probability measure $f_j \in \Delta(\mathbb{N})$ specifying a distribution over the number of workers to hire for j, where $\Delta(\mathbb{N})$ is the set of all probability measures over the nonnegative integers. Let $q_j = \sum_{n=2}^{\infty} f_j(n)$ be the probability that multiple workers are hired for task j and $n_j \in \mathbb{N}$ be the realized number of workers hired for task j. Once the number of workers has been decided upon for a task, each worker-task is assigned to the next available worker. The set $\mathcal{I}_j$ represents the set of workers hired for task j.

The worker-task: A worker i assigned to task j chooses whether to exert high effort ($e_j^i = 1$), in which case he perfectly observes the state of the world and incurs costs c, or low effort ($e_j^i = 0$), which yields no information about the state and is costless. Let $s_j^i \in \{\Omega \cup \emptyset\}$ be the information worker i observes about task j. Effort choices are *unobservable* by the firm and information is *not verifiable*. After making an effort choice and observing information about the state, the worker sends a message to the firm, $m_j^i \in \Omega$.[13]

Back to the task: Upon receiving messages from all hired workers on a task, the firm compiles a message profile,

$$m_j = \{m_j^i\}_{i \in \mathcal{I}_j} \in \mathcal{M}_j$$

to inform its action choice, where $\mathcal{M}_j = \cup_{n=1}^{\infty} \Omega^n$ is the set of all possible task message profiles. The firm's strategy is a mapping from this message profile to the set of probability measures over its action space,

$$\alpha_j : \mathcal{M}_j \to \Delta(\Omega)$$

We assume that a firm's strategy is independent across tasks.

The Job

A job $\mathcal{J}^i$ consists of the set of tasks assigned to worker i, where $J^i = |\mathcal{J}^i|$ denotes the size of the job. We restrict the maximum job size the firm can create to be $\overline{J} < \infty$ to capture exogenous legal or technological constraints that the firm may face.[14] Once a worker receives his job, he completes each worker-task and sends a message profile

$$m_i = \{m_j^i\}_{j \in \mathcal{J}^i} \ \in \mathcal{M}^i$$

to the firm, where $\mathcal{M}^i = \Omega^{J^i}$ is the set of all possible job message profiles.

The worker's strategy specifies an effort level and a message for each task in his job. The effort strategy specifies the probability of exerting effort on each task,

$$\sigma^i = (p_1^i, ..., p_{J^i}^i) \in [0,1]^{J^i}$$

where $p_j^i = Pr(e_j^i = 1) \in [0,1]$ is the probability worker i exerts high effort on task j. Define $\overline{\sigma} = (\overline{\sigma}^i, \overline{\sigma}^{-i})$ and $\underline{\sigma} = (\underline{\sigma}^i, \underline{\sigma}^{-i})$ as the strategy profiles for always working and always shirking, respectively. The message strategy specifies a message profile as a function of the information observed about all tasks[15]

$$\rho^i : (\Omega \cup \emptyset)^{J^i} \to \Delta \mathcal{M}^i$$

[13] With limited liability, it is without loss of generality to restrict the message space to Ω.

[14] Workers may not complete multiple jobs, so there is no scope for repeated interaction.

[15] The information partitions the effort choice of the worker, so it is without loss of generality to define the message as a function of only information.

Let $\overline{\rho}^i$ denote the strategy where (1) the agents reports signals truthfully for tasks on which he exerts effort and (2) the agent reports the state with a higher prior for tasks on which no effort is exerted.

Contracts

So far, we have discussed how a firm chooses the number of workers to hire for each task, and the number of tasks to assign to each worker (the job). These are the first two components of a worker's contract; the third component is the payment scheme.

Effort is unobservable and information is not verifiable; therefore, the payment scheme can only condition on the message profile of a worker, ρ_i, and the messages of the other workers assigned to each task in the worker's job. The firm constructs a report for worker i:

$$r^i = \bigcup_{j \in \mathcal{J}^i} m_j = \bigcup_{j \in \mathcal{J}^i} \bigcup_{i \in \mathcal{I}^j} m_j^i \ \in \mathcal{R}^i$$

where $\mathcal{R}^i = \mathcal{M}^j \times ... \times \mathcal{M}^j$.[16] The firm offers a payment scheme as a function of this constructed report. Let

$$T^i(r^i) : \mathcal{R}^i \to \mathbb{R}_+$$

specify the payment obtained by worker i for job $\mathcal{J}^i$ when report r^i is received. We assume that workers are protected by limited liability, and restrict attention to payment schemes $T^i(r^i) \geq 0$ for all r^i. It will often be convenient to work with the per-task wage $w^i(r^i) = \frac{T^i(r^i)}{J^i}$.

Define an implementation plan for worker i as the collection of probability measures over the number of workers hired for each task in worker i's job,

$$Q^i = \{f_j\}_{j \in \mathcal{J}^i}$$

Then we can formally represent a contract for worker i as a set of tasks, an implementation plan and a payment scheme,

$$\mathcal{C}^i = \left(\mathcal{J}^i, Q^i, T^i(\cdot)\right)$$

It is useful to partition $\mathcal{R}^i$ into subsets $\{\mathcal{R}^i_{kj}\}_{k=0,...,J,j \leq k}$. $\mathcal{R}^i_{kj}$ is the collection of reports from worker i where additional agents were hired for k tasks and all workers make the same recommendation on $j \leq k$ of these tasks; define $\overline{\mathcal{R}}^i = \bigcup_{k=0}^{J} \mathcal{R}^i_{kk}$.

Payoffs

A worker's payoff for a job i depends on the transfer received from the firm and the cost of effort. We assume workers are risk neutral, so a worker's expected utility, given a contract $\mathcal{C}^i$ and strategy profile (σ, ρ), is represented as:

$$U_i(\sigma, \rho; \mathcal{C}^i) = E\left[T^i(r^i) - c \sum_{j=0}^{J^i} 1_{\{e^i_j = 1\}}\right]$$

[16] Reports for different workers will contain common elements when the workers have overlapping tasks.

Note that the worker's payoff is independent of the firm's strategy.

The risk-neutral firm seeks to match the realized state of the world for each task. The firm must pay each agent the specified transfer for their employment. The firm's expected flow payoffs from an incentive contract for i, given worker strategy profile (σ, ρ), firm strategy profile α and contract $\mathcal{C}^i$, is represented as:

$$V(\alpha, \sigma, \rho; \mathcal{C}^i) = E\left[\sum_{j=1}^{J} \frac{\mathbf{1_{A_j=\omega}}}{|\mathcal{I}_j|} - T^i(r^i)\right]$$

The general problem

The firm maximizes its expected flow payoffs from an incentive contract by choosing a job size, monitoring technology, transfer scheme, and action profile for each possible report, all subject to the strategy profile of the hired workers being optimal. That is, the firm chooses $(\mathcal{C}^i, \alpha)$ such that:

$$\max_{\mathcal{C}^i, \alpha} V(\alpha, \sigma, \rho; \mathcal{C}^i) \tag{*}$$

subject to

$$(\sigma^i, \rho^i) \in \arg\max_{(\hat{\sigma}^i, \hat{\rho}^i)} U_i(\sigma, \rho; \mathcal{C}^i) \quad \forall i \tag{IC}$$

$$U_i(\sigma, \rho; \mathcal{C}^i) \geq 0 \quad \forall i \tag{IR}$$

$$T^i(r^i) \geq 0, \quad \forall i,\ r^i \in \mathcal{R}^i \tag{LL}$$

$$E[1_{A_j=\omega}] \geq \pi \quad \forall j \tag{FP}$$

The maximand of * is the firm's objective function; the remaining expressions capture worker incentive compatibility and individual rationality, limited liability, and firm participation. FP states that the firm must prefer the equilibrium to not entering the contractual relationship.

Section 5 considers the class of contracts for which the firm lacks incentive for induced breach. In other words, a report without evidence of shirking is deemed acceptable for the purposes of wage payments. In particular, the firm's failure to monitor workers is not just cause for withholding payments. The no-induced breach condition is stated as B.[17]

$$T^i(r^i) = t \ \forall r^i \in \overline{\mathcal{R}}^i. \tag{B}$$

It is worth noting that B places no restrictions on firm behavior when there is evidence of shirking.

Definition 1 (Equilibrium)**.** *An organizational equilibrium (henceforth equilibrium) is a solution to the firm's problem satisfying B in which workers exert high effort and report truthfully.*

The problem facing the firm remains quite general: the firm is able to stochastically hire workers for each task and offer non-linear payment schemes to reward the agents. Equilibrium continues to take a simple structure and, as Section 5 demonstrates, the efficiency loss relative to contractible effort vanishes as jobs grow large.

5 Analysis

It is useful to begin with the first-best equilibrium when effort is contractible. The equilibrium must only satisfy the participation constraints of the worker and firm (IR and FP). Worker participation is ensured by $w \geq c$. The firm's equilibrium payoff for each task is $1 - c$, which it prefers to guessing the state blindly for $c \leq 1 - \pi$.

Remark 1 (Contractible Effort Benchmark)**.** *When effort is contractible and $c \leq 1 - \pi$, Q^* is such that $q_j = 0$ for all j and*

$$T^*(r^i) = \begin{cases} J \cdot c & \forall\ r^i \ \in \overline{\mathcal{R}}^i \text{ and } (\sigma^{\mathrm{i}}, \rho^{\mathrm{i}}) = (\overline{\sigma}^{\mathrm{i}}, \overline{\rho}^{\mathrm{i}}) \\ 0 & \forall\ r^i \ \in \overline{\mathcal{R}}^i \text{ and } (\sigma^{\mathrm{i}}, \rho^{\mathrm{i}}) \neq (\overline{\sigma}^{\mathrm{i}}, \overline{\rho}^{\mathrm{i}}). \end{cases}$$

The contractible effort benchmark is unattainable here and, more generally, there is no equilibrium with $q_j = 0$ for any task: since a worker's report reveals nothing in isolation about his effort choice, the firm cannot ascertain effort and workers will shirk upon accepting such a contract. The lesson is that the firm must hire additional workers to induce effort. The workers duplicate tasks, enabling the firm to compare output across agents. The firm generates incentives for effort by conditioning payment upon reports it deems acceptable.

Monitoring leads to unavoidable redundancy as multiple people are dedicated to the same project. There is no distinction between supervisors and subordinates in the model. Instead,

---

[17] In the optimal equilibrium without B, the firm pays a worker if and only if multiple workers are employed on each of the worker's tasks and the output they produce always coincides. (See Result 3 in Appendix B.) The contracts used in practice bear little similarity to this bounty-like structure (Dickens, Katz, Lang, and Summers, 1989).

Limited liability alone does not sufficiently constrain the firm's behavior, suggesting a relevant feature of the firm's problem is omitted. The contract treats as justification for nonpayment both evidence of shirking and a lack of evidence of working. This is one reason why adhesion contracts, like those considered here, traditionally minimize the actionable obligations of the firm to the payment of money. Posner (1972) and Farber (1980) note that courts will not enforce supercompensatory damages caused by a breach of contract. This reluctance is explained by a desire to avoid creating incentives for "induced breach."

the firm must create a monitoring apparatus by hiring more than one worker for each task and having the workers serve as monitors for each other.

The analysis follows by dividing the principal's problem into two stages: first, a job size J is given and for any monitoring technology Q^i there is shown to be another monitoring technology with constant monitoring probability that the firm prefers. The optimal wage given this superior monitoring technology is then derived. Theorem 1 therefore describes necessary properties of the optimal incentive organization. The optimal monitoring technology is then determined by maximizing over all technologies and job sizes satisfying Theorem 1.

5.1 Implementation

Lemma 1 simplifies the exposition by establishing two results. First, the firm never hires more than two workers for any task in equilibrium so it is without loss of generality to describe implementation plans as $Q^i = (q_1, ..., q_J)$ instead of $Q^i = (f_1, ..., f_J)$. Second, it is without loss to restrict attention to truthful recommendation strategies $\overline{\rho}$ in which workers report their signals truthfully and follow the prior when shirking. Recommendation strategies are omitted in the discussion that follows and attention is focused on inducing effort.

Lemma 1. *For any equilibrium of the firm's problem, there exists another equilibrium providing the same expected payoffs to the firm and all agents in which (a) the firm adopts an implementation plan $Q^i = (q_1, ..., q_J)$ with $q_j = \sum_{n=2}^{\infty} f_j(n)$ and $f_j(n) = 0$ for all $n > 2$, and (b) workers use truthful recommendation strategies $\rho = \overline{\rho}$.*

The firm wants to induce effort at the lowest expected cost. One worker is insufficient and hiring two workers deterministically is costly. The firm may be able to improve its payoffs by stochastically employing a second worker.

The firm needs to align worker incentives so that high effort is an equilibrium. Consider generic implementation plan $Q^i = (q_1^i, ..., q_J^i)$ and strategy profile $(\sigma^i, \overline{\sigma}^{-i})$. The implementation plan and the strategy profile induce a distribution on the partition $\{\mathcal{R}_{kj}^i\}_{k=0,...,J,j\leq k}$. Let $Pr(\mathcal{R}_{kj}^i) = \sum_{r^i \in \mathcal{R}_{kj}^i} G(r^i|\sigma^i, \overline{\sigma}^{-i}, Q^i)$ be the probability of a report in partition element $\mathcal{R}_{kj}^i$ when worker i is following strategy σ^i and all other workers are exerting effort. The induced distribution on this partition is represented in Figure 2. Similarly define $\overline{Pr}(\mathcal{R}_{kj}^i) = \sum_{r^i \in \mathcal{R}_{kj}^i} G(r^i|\overline{\sigma}^i, \overline{\sigma}^{-i}, Q^i)$.

The firm's payment scheme details lump-sum payments as a function of the report. Denote by $t_{kj} = T^i(r^i)$ for $r^i \in \mathcal{R}_{kj}^i$ as the transfer to worker i when a monitor is employed on $k \leq J$ tasks and output matches on $j \leq k$ of these tasks.

The general incentive constraint can now be written. Incentive compatibility requires

$$\begin{aligned} & t_{00}\overline{Pr}(\mathcal{R}_{00}^i) + t_{10}\overline{Pr}(\mathcal{R}_{10}^i) + ... + t_{JJ}\overline{Pr}(\mathcal{R}_{JJ}^i) - Jc \\ & \geq t_{00}Pr(\mathcal{R}_{00}^i) + t_{10}Pr(\mathcal{R}_{10}^i) + ... + t_{JJ}Pr(\mathcal{R}_{JJ}^i) - c \cdot \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i) \end{aligned} \tag{1}$$

for all strategies σ^i.

Regardless of the transfer scheme, hired workers can choose to shirk and report to the firm *as if* they exerted effort and acquired signals. There is always positive probability

Monitored \ Match	0	1	...	J
0	$\Pr(\mathcal{R}^i_{00})$			
1	$\Pr(\mathcal{R}^i_{10})$	$\Pr(\mathcal{R}^i_{11})$		
...	...	...	...	
J	$\Pr(\mathcal{R}^i_{J0})$	$\Pr(\mathcal{R}^i_{J1})$	...	$\Pr(\mathcal{R}^i_{JJ})$

Figure 2: Distribution on $\{\mathcal{R}^i_{kj}\}_{k=0,\dots,J,j\le k}$ induced by $(\sigma^i, \overline{\sigma}^{-i})$ and Q^i.

that such a strategy garners payment and so satisfaction of incentive compatibility implies satisfaction of individual rationality. That IR holds strictly implies workers obtains rents, even when the firm structures incentives optimally.

The role of limited liability can be seen by considering Equation 1 without imposing LL. Unlimited liability allows the firm to threaten workers with arbitrarily severe punishments. To dissuade σ^i, the firm need only set t_{kj} negative enough for $r^i \in \mathcal{R}^i_{kj}$ such that $\overline{Pr}(\mathcal{R}^i_{kj}) < Pr(\mathcal{R}^i_{kj})$.

With limited liability $t_{kj} \ge 0$ and such threats are unavailable. The firm must dissuade shirking by providing workers with rents in equilibrium. As in Shapiro and Stiglitz (1984), threats take the form of losing these rents if caught deviating.

The incentive constraint can be simplified by noting the firm should never offer a positive transfer when a worker fails to match on any of his monitored tasks. Setting $t_{kj} > 0$ for $k \ne j$ is both directly and indirectly costly: not only are workers paid when reports in $\mathcal{R}^i_{kj}$ are realized, but $t_{kj} > 0$ makes Equation 1 harder to satisfy since $\overline{Pr}(\mathcal{R}^i_{kj}) \le Pr(\mathcal{R}^i_{kj})$ whenever $k \ne j$.

Lemma 2. *The optimal incentive contract sets* $T^i(r^i) = 0 \ \forall r^i \in \mathcal{R}^i_{kj}$ *whenever* $k \ne j$.

Lemma 2 implies workers are required to produce matching output whenever multiple workers are assigned to a task. Each worker's transfer is therefore a function of the reports of all workers. Even though workers produce output individually, the payment scheme treats them *as if* they are a team and punishes everyone when any one worker shirks.

From $\sum_{k=0}^{J} \overline{Pr}(\mathcal{R}^i_{kk}) = \overline{Pr}(\overline{\mathcal{R}}^i) = 1$ and the no induced breach condition (B), Equation 1 can be rewritten in terms of the required transfer t:

$$t \ge \frac{c(J - \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i))}{1 - Pr(\overline{\mathcal{R}}^i)}, \quad \forall \sigma^i \in \Sigma^i. \tag{2}$$

The incentive constraint is further simplified by viewing the firm's zero tolerance of mismatches as the firm requiring acceptable output on every assigned task. Acceptable is then taken to include the vacuous case in which only one worker was assigned the task.

Worker incentives are governed by the probability a worker believes his report will be compared to the report of another agent. So while implementation plan $Q^i = (q_1, ..., q_J)$ specifies hiring two workers for task j with probability q_j, a worker assigned j believes with probability $\frac{1-q_j}{1+q_j}$ he is the only agent employed. With probability $\frac{2q_j}{1+q_j}$ the worker is one of two agents assigned j.

Generic strategy $\sigma^i = (p_1^i, ..., p_1^J)$ specifies the probability i exerts effort on each task $j \in \mathcal{J}^i$. The worker produces acceptable output on j with probability $\frac{1-q_j+2q_j(p_j^i+(1-p_j^i)\pi)}{1+q_j}$. Since the firm requires acceptable output on all assigned tasks, the incentive compatible wage satisfies

$$t \geq \frac{c(J - \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i))}{1 - \left[\frac{1-q_j+2q_j(p_j^i+(1-p_j^i)\pi)}{1+q_j}\right]^J}, \quad \forall \sigma^i \in \Sigma^i. \tag{3}$$

The payment required to induce effort can now be derived. Theorem 1 establishes several properties of an optimal contract. It shows that implementation plans employing different monitoring probabilities make inefficient use of the firm's monitoring ability. It also shows that it is sufficient for the firm to dissuade $\underline{\sigma}^i$.

Theorem 1. *The optimal contract will take the form of an implementation plan $Q^i = (q, ..., q)$ that specifies hiring a second agent with probability q for each task and a transfer*

$$T^i(r^i) = \begin{cases} J \cdot w^*(q, \pi, c, J) & \forall\ r^i\ \in \overline{\mathcal{R}}^i \\ 0 & \forall\ r^i\ \notin \overline{\mathcal{R}}^i, \end{cases}$$

where the equivalent per-task wage is $w^*(q, \pi, c, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$.

Corollary 1 (Comparative Statics)**.** *$w^*(q, \pi, c, J)$ is decreasing in q and J and increasing in π and c.*

Theorem 1 establishes a schedule of implementation plans and transfer schemes capable of inducing equilibrium behavior. The (Q^i, T^i) combinations trade-off higher monitoring probabilities against lower wages.

As in the contractible effort benchmark, the firm retains the possibility of guessing the state blindly instead of hiring workers. Theorem 1 implies there exist effort costs for which the firm would undertake an employment contract were effort contractible but not when effort is non-contractible.

Corollary 2. *For effort costs $c \in ((1-\pi)\frac{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}{1+q}, 1-\pi]$ the firm will not employ any workers despite it being efficient to do so were shirking not a concern.*

Theorem 1 relies on two main results. First, since a worker is maximizing a linear objective function over a convex set, the most attractive deviation away from always working is to a deterministic effort strategy. Second, since the probability of producing matching output when shirking on any task is log-convex, the firm optimally chooses an implementation plan specifying a common monitoring probability for all tasks. This is because an implementation

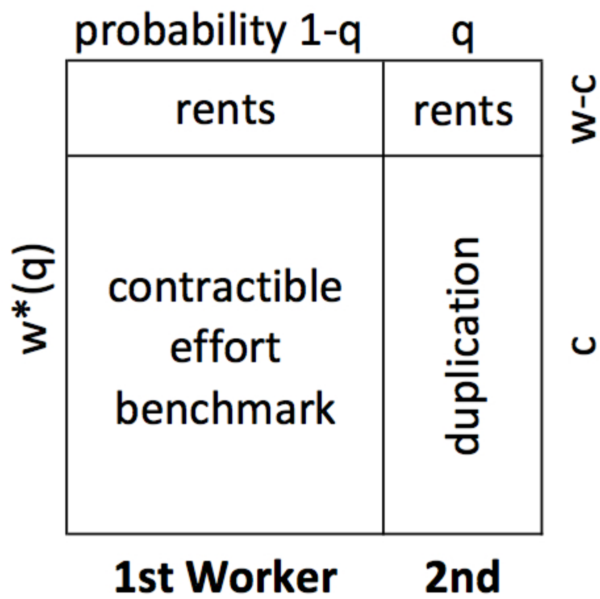

Figure 3: Composition of per-task wage bill.

plan with common monitoring probability $\overline{q} = \frac{1}{J}\sum_{k=1}^{J} q_k$ specifies a lower wage paid to the same expected number of workers as $Q^i = (q_1, ..., q_J)$. Alternately, the firm can adopt a common monitoring probability $\hat{q}$ where $\left[\frac{1-\hat{q}+2\pi\hat{q}}{1+\hat{q}}\right]^J = \left[\frac{1-q_1+2\pi q_1}{1+q_1}\right] \cdot ... \cdot \left[\frac{1-q_J+2\pi q_J}{1+q_J}\right]$. This implementation plan pays the same wage to fewer expected workers.

Lemma 4 establishes that given any implementation plan $Q^i = (q_1, ..., q_J)$, there exists another implementation plan $\hat{Q}^i = (q, ..., q)$ that offers the same transfer scheme but hires fewer workers. Optimality is then a case of trading off the stronger disincentives (and lower wages) of more intensive monitoring against the greater number of expected workers.

5.2 Optimization

Theorem 1 begins by considering any implementation plan Q^i for a J-sized job. It shows that there exists another implementation plan specifying a constant monitoring probability across tasks that the firm prefers to Q^i. Theorem 1 then derives the minimum transfer required to induce effort. Thus, Theorem 1 describes several properties the optimal organization of incentives must satisfy.

The monitoring probability q is now taken to be endogenously chosen by the firm. In addition to being a key component of the firm's transfer scheme, the implementation plan also affects how many workers the firm expects to hire. The size of the job, J, continues to be exogenously given for the time being.

The firm learns the state for each task in equilibrium. Since any (Q^i, T^i) combination satisfying Theorem 1 provides the same information about the unknown states to the firm, optimality is determined by minimizing the firm's expected wage bill.

With monitoring probability q, $1 + q$ workers are hired in expectation for each task

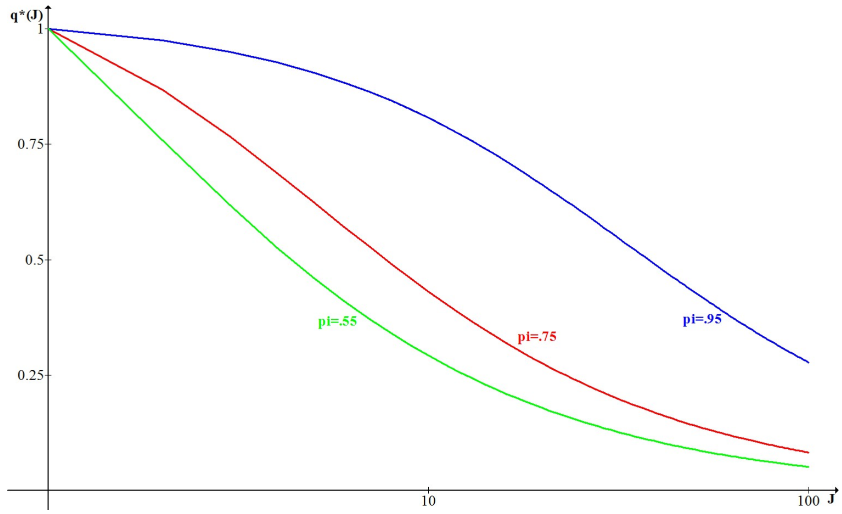

Figure 4: Optimal monitoring probability as a function of the job size.

and each is paid the equivalent of $w^*(q, \pi, c, J)$ per task.[18] The firm's optimal monitoring probability then solves

$$q^*(\pi, c, J) = \arg\max_q \ (1+q) \cdot \frac{c}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J}. \tag{4}$$

The optimal monitoring probability is implicitly determined to balance the cost of greater monitoring – more workers are hired in expectation – against the benefit of lower wages. While $q^*(\pi, c, 1) = 1$, for $J > 1$ $q^*(\pi, c, J) < 1$ and the firm does not hire a monitor deterministically.

Theorem 2 (Stochastic Monitoring)**.** $q^*(\pi, c, J) < 1$ *if and only if* $J > 1$.

The optimal monitoring probability responds intuitively to changes in the environment. (See Figure 4.) As the prior becomes more pronounced, shirking becomes more attractive and the firm must monitoring employees more frequently. The optimal monitoring probability does not vary with changes in the cost of effort; instead, the firm adjusts the transfer.

From the firm's perspective, monitoring and bundling are substitutes and the firm trades the intensive margin for the extensive margin: each task is monitored less intensively in exchange for assigning more tasks.

Theorem 3. $q^*(\pi, c, J) \to 0$ *as* $J \to \infty$.

Theorem 1 shows that for any job size the firm can approximate the action profile of the contractible effort benchmark. Theorem 2 shows that this is achieved at great cost: the firm can obtain higher expected payoffs by monitoring more often and not approximating the benchmark action profile. While Rahman (2012) considers when approximating the contractible effort benchmark action profile is feasible, Corollary 3 shows that this is optimal only in the limit.

[18]Implicit in writing the per-task wage bill as $(1 + q) \cdot w$ is that when two workers are hired for a task, both are paid the same wage. This is shown as a property of the optimum in Section 6.1.

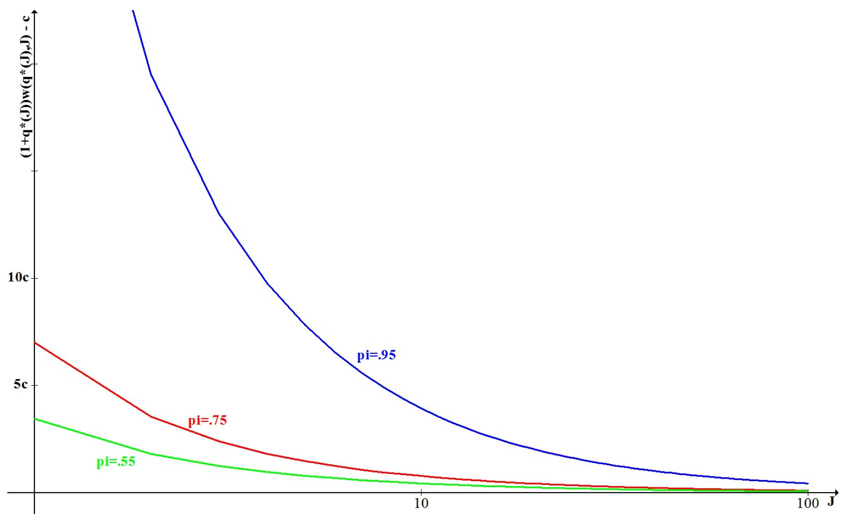

Figure 5: Efficiency loss as a function of the job size.

Corollary 3. *For any (π, c), given $\varepsilon > 0$, $\exists\ J_\varepsilon < \infty$ such that $q^*(\pi, c, J) > \varepsilon$ for all $J < J_\varepsilon$.*

Remark 2 (Virtual Monitoring Only in Limit)**.** *When monitoring is chosen endogenously, virtual monitoring is adopted only in the limit as $J \to \infty$.*

For fixed J, Corollary 1 shows $w^*(q, \pi, c, J)$ is increasing in q. One may suspect $w^*(q^*(\pi, c, J), \pi, c, J)$ to be increasing in J; after all, if the firm is rarely employing a monitor, the worker has a high probability of producing acceptable output and so requires a higher wage to eschew shirking. This fails to account for the direct effect on w^* of increasing J, however: shirking is less attractive for large J since the worker needs to produce acceptable output for more tasks. Along the optimal path as J increases, the direct effect of increasing J on $w^*(q^*(\pi, c, J), \pi, c, J)$ dominates the indirect effect of decreasing q^*. Not only are fewer workers being hired for each task, but each worker is paid a lower per-task wage. Taken together, the efficiency loss to the firm relative to the contractible effort benchmark vanishes asymptotically as $J \to \infty$. (See Figure 5.)

Result 1 (Asymptotic Efficiency)**.** *As $J \to \infty$, $(1 + q^*(\pi, c, J)) \cdot w^*(q^*(\pi, c, J), \pi, c, J) \to c$.*

The implementation and surplus extraction literature surveyed in Section 2 treats the principal-agent problem as a problem of inserting a wedge between working and shirking. When negative payments are possible, the wedge can be monetary punishments for poor (or unfortunate) outcomes. With limited liability, sticks are unavailable and a carrot must be used. The wedge becomes a reward that is withheld for poor performance.

Bundling presents the firm another option when dismissal and punishment are unavailable. Instead of a costly bonus on top of the worker's earnings, the reward is simply the receipt of the wages the worker accumulated throughout the job. By bundling multiple tasks together into one job, the firm is able to withhold the worker's earnings until the entire job is successfully completed. The disincentive to shirking would be dulled if the firm instead paid on a piece-rate basis.

Remark 3. *The firm optimally bundles as many tasks together as possible: $J^* = \overline{J}$.*

6 Discussion

Bundling takes the form of assigning a worker multiple tasks and conditioning payment for any task on satisfactory performance on all tasks. An alternative interpretation of bundling is provided by considering multidimensional tasks.

Suppose an agent is assigned a chore consisting of multiple components. For example, a worker must complete a tax return with additional schedules for each source of non-wage income. The tasks of Section 4 are now individual tax forms and a job is the entire tax return. The main result of Section 5 is that the Internal Revenue Service should monitor individual tax forms and impose the harshest possible penalty upon uncovering any irregularities. There is evidence the IRS follows such a strategy. Tax returns claiming unusually large charitable deductions, Schedule C self-employment income or business expenses invite an audit of the entire return (Barrett, 2011; CBSNews, 2010).

An entertaining example of the theory of Section 5 is provided by the rock band Van Halen. Like many musical acts, the band's performance contract with event venues is a long, complicated document specifying hundreds of individual items. Within the 53-page rider is an obscure provision often taken as prima facie evidence of rock excess: a bowl of M&M's is to be provided with all brown candies removed. As the band's lead signer explained in his autobiography, the unusual request performed a monitoring function:

> Van Halen was the first band to take huge productions into tertiary, third-level markets. We'd pull up with nine eighteen-wheeler trucks, full of gear, where the standard was three trucks, max. And there were many, many technical errors – whether it was the girders couldn't support the weight, or the flooring would sink in, or the doors weren't big enough to move the gear through.
>
> The contract rider read like a version of the Chinese Yellow Pages because there was so much equipment, and so many human beings to make it function. So just as a little test, in the technical aspect of the rider, it would say "Article 148: There will be fifteen amperage voltage sockets at twenty-foot spaces, evenly, providing nineteen amperes ..." This kind of thing. And article number 126, in the middle of nowhere, was: "There will be no brown M&M's in the backstage area, upon pain of forfeiture of the show, with full compensation."
>
> So, when I would walk backstage, if I saw a brown M&M in that bowl ... well, line-check the entire production (Roth, 1997, pp. 97-98).

Such an arrangement is not unusual. Managers hiring subordinates are frequently faced with the problem of judging a worker's ability on the basis of limited information. As anyone with experience in such hiring decisions can attest, candidates are often removed from consideration for having spelling or grammar mistakes on their resumes. Poor performance on one component is taken as evidence of low quality on other, unrelated dimensions. For the same reason an investing website lists "Math Errors" and "Failure to Sign the Return" as the numbers 2 and 3 "red flags" prompting a tax audit (Investopedia, 2010).[19]

Allen (2011, p. 43) suggests an inability to monitor individual tasks led to bundling before the industrial revolution: "Paying workers for specific tasks meant both parties needed to be

[19] The number 1 "red flag" is "Overestimating Donated Amounts."

able to separate these tasks from others. The inability to do this during the pre-modern era meant that all tasks were essentially bundled together, and as a result, the laborer became a servant under the general and universal supervision of the master." The argument put forth in this paper is that bundling becomes even more attractive when individual measurement is possible since it permits the principal to strengthen the incentives provided to agents.

The incentive structure of Section 5 is similar to that employed by the British Admiralty. The navy struggled with cowardice and needed to arrange adequate rewards supported by an effective monitoring apparatus. Opportunities to amass wealth were made hostage to acceptable performance. "Any slipup discovered by the Admiralty meant [...] that half pay was given to those captains who made slight mistakes" (Allen, 2011, p. 123).

Monitoring was performed by the battle line and multiple record keeping. With the creation of Fighting Instructions in the 1600s, the navy codified a set of rules to enable easier identification of shirking at sea. The instructions specified the formation of a "line of battle" with all ships keeping in line with the chief. The tactical disadvantage of the battle line was outweighed by the ease of monitoring it offered: identifying when a captain was failing to engage in battle was simply a matter of spotting who was "out of line."

With the creation of the battle line, the Admiralty invested in a costly monitoring technology to generate desirable behavior in battle. To induce desirable behavior outside of battle, the Admiralty required commissioned lieutenants and noncommissioned masters on board each ship to keep detailed journals of the captain's performance. The journals were turned over to the Admiralty upon reaching shore and different accounts of the officers' performance were compared.

Costly duplication – like requiring multiple overlapping accounts of a captain's actions – is a feature common throughout the experimentation literature. There, additional signals are acquired for learning purposes. Each signal increases the firm's knowledge about an unknown state and the firm acquires signals until the marginal benefit from better information is outweighed by the marginal cost of purchasing another signal.

There is no learning justification for hiring multiple workers in the firm's problem in Section 4. Hiring multiple agents is for incentive reasons only. Any departure from the contractible effort benchmark is the cost of generating adequate incentives to induce effort. The theory thus provides a complementary rationale for costly duplication to that offered in the experimentation literature. In the example of the British Admiralty, both motivations were likely in play, though the historical record suggests monitoring was the larger concern.

6.1 Hierarchy

Alchian and Demsetz (1972) suggest the firm best aligns incentive by having specialized monitors. It may seem reasonable to suspect such a result here: a worker believing with certainty that two workers are assigned to a task knows his output will be verified and, thus, requires a lower wage in equilibrium. The model of Section 4 can be modified to examine this question.

It is assumed in Section 4 that workers know only the firm's implementation plan for each assigned task. That is, workers know how many workers the firm expects to hire for each task, but a hired worker does not know whether he is the first or second agent the firm hires for the task. Each hired worker believes a task is assigned to two agents with probability

$\frac{2q}{1+q}$. This need not be the case. Instead, the firm can inform each worker it hires whether it is the first agent assigned the task or the second agent. The firm creates specialized monitors by doing so.

Given any monitoring probability q, $1+q$ workers are hired in expectation on each task. The probability that a given worker is the first agent assigned to a task is $\frac{1}{1+q}$; likewise, with probability $\frac{q}{1-q}$, a randomly chosen worker is the second agent assigned to the task.

A worker informed that he is the first agent hired knows his output is monitored with probability q. A worker told that he is the second agent assigned to a task knows that his output is necessarily checked by another agent. Thus, before informing a worker about whether he is the first or second agent hired, the worker's ex ante belief about the probability his output is monitored is given by $\frac{1}{1+q}q + \frac{q}{1+q}1 = \frac{2q}{1+q}$. This implies that dedicating certain workers as specialized monitors results in a mean-preserving spread of beliefs when q is known by all workers. Since Theorem 1 implies that the wage required to induce effort is convex in a worker's belief that a task is assigned to two agents, the firm's expected wage bill is minimized by treating workers symmetrically.

Remark 4 (Hierarchy)**.** *The optimal organization of the firm does not include a hierarchy of monitors and subordinates.*

Contrary to the suggestion of Alchian and Demsetz (1972), creating a vertical hierarchy within the firm is not optimal.[20] While it is true that subordinates – those who believe their work is often verified by a supervisor – receive lower wages within a hierarchical organization, supervisors believe their output is often unverified by the firm and therefore command a higher wage. In the context of the model of Section 4, supervisors are akin to the first worker employed on a task: only sometimes is their output is compared to that of another agent. Subordinates are like the second worker assigned to a task. By definition first workers are more numerous than second workers.

Alchian and Demsetz (1972, p. 782) famously ask "Who will monitor the monitor?" The answer provided here is "the monitored": the monitor of the monitor is the monitored. The dynamic of workers simultaneously monitoring each other is a key feature of Section 5.

6.2 Virtual monitoring

An incentive contract exhibits virtual monitoring if for any $\varepsilon > 0$ the firm induces a high-effort equilibrium while hiring a monitor with probability less than ε. In the contractible effort benchmark (Remark 1), each worker exerts effort and the firm never employs a monitor. Interest in virtual monitoring arrangements stems from the action profile under virtual monitoring being arbitrarily close to the action profile under the contractible effort benchmark.

Rahman (2012) examines when virtual monitoring is feasible for the firm. From Theorem 1, virtual monitoring is always feasible but it requires unboundedly large wage payments. While the firm can approximate the first-best action profile, it cannot approximate its payoffs in the first-best equilibrium. Since the required wage diverges to infinity as the monitoring probability shrinks, virtual monitoring is not a feature of the optimal incentive organization.

[20] Williamson (1967), Mirrlees (1976) and Calvo and Wellisz (1978) also discuss firm size.

Remark 5 (Virtual Monitoring)**.** *There exists a transfer scheme capable of inducing equilibrium for any monitoring probability* $q > 0$*, but* $w^*(q, \pi, c, J) \to \infty$ *as* $q \to 0$*.*

6.3 Mediated contracts and the "Gold Standard"

In a mediated contract, a third-party, or mediator, provides agent-specific action recommendations to introduce correlation among agents' action choices (Rahman and Obara, 2010). Rahman (2012) studies a multilateral principal-agent problem with costly effort where payments depend on the recommendation-contingent actions of the workers.

By privately recommending actions to workers, the firm asks agents a question for which it already knows the answer. In Rahman (2012), the question posed to a monitor is whether a subordinate exerted effort; the firm randomly instructs the subordinate to shirk and rewards the monitor only for correctly reporting the subordinate's action.

Such mediated contracts are not literally available in our setting. However, the dynamic of posing workers with questions for which the firm already knows the answer can be recreated just the same. In fact, the use of questions with known answers is already commonplace in crowdsourcing, where it's known as the "gold standard."[21] The firm simply seeds each worker's job with tasks for which the state has already been learned. Workers are paid if they perform satisfactorily on this subset of tasks.

Regardless of their availability, these mediated-style contracts are not optimal here. To see this, suppose that the firm has costlessly acquired a collection of tasks for which the state is privately known to the firm. In actuality, the firm would have had to hire workers to discover the state, thereby making this knowledge costly to acquire in the first place, but the suboptimality of an incentive organization built around known tasks can be shown even in this more demanding setting.

Workers are assigned J tasks and the firm optimizes over how many known tasks to include within this set; let n be the optimal number. Worker reports on their J-sized jobs are judged on the basis of their performance on the subset of known tasks. Payment is provided if and only if the worker performs satisfactorily on all n tasks.

Assigned J tasks, a worker must be compensated for exerting effort on all J tasks. (Workers must also receive rents to dissuade them from shirking, as above.) Call this compensation $\tilde{w}$. The firm is paying each worker $\tilde{w}$ in equilibrium and is obtaining knowledge of the state on $J - n$ new tasks. So with $2J$ tasks assigned to two workers, the cost is $2\tilde{w}$ while the benefit is $2(J - n)$.

Alternatively, the firm could tell each worker that he will be monitored on n of the J tasks the firm assigns him. The disincentive to shirking is just the same, so two workers will cost the firm just as much as before, $2\tilde{w}$. But now the firm is learning the state for $J + (J - n) > 2(J - n)$ new tasks. The same wage bill is now spread over a greater number of valuable tasks: the effective per-task wage is reduced by using an incentive organization based around the results in Section 5. Mediated-style contracts are inefficient because they monitor workers independently; instead, the optimal incentive organization in Section 5 monitors workers simultaneously. This advantage is reduced as J grows large.

[21] See www.crowdflower.com, for example.

Remark 6. *Mediated-style contracts, like the "Gold Standard," are dominated by the incentive organization of Section 5 that monitors multiple workers simultaneously.*

6.4 Robustness

Section 4 presents a benchmark model in which effort perfectly reveals a task's unknown state and all workers share a binary cost of effort. This stark setting best isolates the incentive issues facing a firm contracting in a labor spot market. The firm hires multiple workers purely for incentive reasons; there is no learning justification for the duplication of tasks.

The main results of Section 5 derive the firm's optimal incentive organization. The firm induces workers to undertake costly effort by periodically monitoring their output through the employment of additional workers. Incentive contracts optimally require workers to complete as many tasks as permissible.

These results obtain outside of the benchmark setting of Section 4. Whether effort costs are a continuously increasing function, workers exhibit heterogenous costs of effort, or exerting effort imperfectly reveals the state, the effect is that mismatches occur in equilibrium and the firm may want to introduce some measure of forgiveness into its incentive structure.

Consider the case in which effort is not perfectly precise. In the benchmark model of Section 4, workers necessarily produce satisfactory output in equilibrium and, therefore, they obtain payment with certainty when working. This is not the case when effort imperfectly reveals the state. Even with all workers exerting effort on all tasks, workers sometimes report different states for the same task and the incentive contracts from Section 5 are unable to induce effort. By displaying leniency upon observing mismatches the firm is able to again convince workers to exert effort.

There are three levers the firm can employ to introduce leniency. It can offer a positive wage payment despite a worker occasionally failing to produce matching output. In addition to setting an interior match rate, the firm can assign fewer tasks to each worker or monitor each task less intensively. The effect of each lever is the same: a worker is able to produce matching output on fewer tasks and still obtain payment. This does not imply the firm is indifferent among the levers though. Indeed, the optimal incentive organization always bundles as many tasks together as possible.

Result 2. *Let $\gamma \in (\pi, 1]$ be the precision of the signal obtained by exerting effort: $\gamma = Pr(s_j = \omega_j | \omega_j)$. The optimal incentive organization bundles as many tasks together as permissible.*

Result 2 establishes that the firm does not introduce leniency by endogenously constraining the size of jobs. So, how does the firm structure incentives? Compared to the benchmark model of Section 4, the firm monitors less frequently when mistakes happen in equilibrium. Instead of directly forgiving mismatches when they occur, forgiveness for mismatches is introduced by the firm detecting mismatches less often. Interestingly, as signals become less precise, the optimal monitoring probability falls. When signals are imprecise mismatches occur more frequently, making the worker's constraints harder to satisfy. The firm relaxes these constraints by monitoring less frequently. Instead of forgiving more mismatches or assigning fewer tasks, the firm elects to catch mismatches less often.

6.5 Other extensions

Generalized payoffs

The firm's payoff is 1 for matching the state and 0 for failing to match the state in Section 4. Allowing the firm's payoffs for matching (and failing to match) the state to depend on the state only affects the firm's participation constraint (FP). Let $\alpha_\omega \geq 0$ be the firm's payoff from matching the state when the state is ω and $\beta_\omega \leq 0$ be the firm's payoff from failing to match state ω. So α_1 (β_1) corresponds to permitting (prohibiting) harmless content to be shared and α_0 (β_0) represents removing (failing to remove) an objectionable item.

Firm optimization remains as before. The sole change is to when the firm prefers the proposed contract to guessing the state blindly. With payoffs α_ω and β_ω, the firm's expected payoff from selecting $A_j = 0$ ($A_j = 1$) without hiring any workers is $(1-\pi)\alpha_0 + \pi\beta_1$ ($\pi\alpha_1 + (1-\pi)\beta_0$). The proposed contract must exceed both of these values, so the firm's participation constraint becomes

$$\pi\alpha_1 + (1-\pi)\alpha_0 - (1+q)\left(\frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}\right) \geq \max\{(1-\pi)\alpha_0 + \pi\beta_1; \pi\alpha_1 + (1-\pi)\beta_0\},$$

which is equivalent to $c \leq \frac{\min\{\pi(\alpha_1-\beta_1);(1-\pi)(\alpha_0-\beta_0)\}}{(1+q)}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right)$. Corollary 2 adjusts accordingly.

Generalized tasks

Section 4 considers a binary state space $\Omega = \{0, 1\}$ to simplify the exposition. It is often without loss of generality to presume tasks can be subdivided into binary choices: instead of asking workers to classify a product as a "white, long-sleeve blouse," the firm can instead structure tasks so that workers separately categorize whether the item is "white," "long-sleeve," and a "blouse."

Even without the possibility of such a subdivision process, the results presented in Section 5 continue to apply with more general state spaces. Let $\Omega = \{\omega^1, \omega^2, ...\}$ and redefine $\pi = \max_k Pr(\omega = \omega^k)$. That is, π is the probability of the most likely state. From the firm's perspective, this is the only state that matters for inducing effort, and so the results in Section 5 carry-through unchanged.

In Section 4 the firm has access to a stream of identical tasks. In particular, $\pi = Pr(\omega_j = 1)\ \forall j$. Suppose instead that for each task j the prior belief is $\pi_j = Pr(\omega_j = 1)$ and that the firm hires an additional worker for j with probability q_j. A worker's probability of matching on task j when shirking is $\frac{1-q_j+2\pi_j q_j}{1+q_j}$.

The incentive compatible wage scheme is determined by modifying Theorem 1. Consider a monitoring technology $(q_1, ..., q_J)$ and $\hat{j} \in \arg\max_{j=1,...,J} \frac{1-q_j+2\pi_j q_j}{1+q_j}$, so that the temptation to shirk is greatest on task $\hat{j}$. If a worker shirks on $\hat{j}$, he shirks on all tasks, so it is sufficient to discourage the worker's strategy calling for no effort.

Now suppose $\frac{1-q_{\hat{j}}+2\pi_{\hat{j}}q_{\hat{j}}}{1+q_{\hat{j}}} > \frac{1-q_j+2\pi_j q_j}{1+q_j}$ for some task j. This cannot be optimal since the firm could lower the monitoring probability on task j – thereby increasing $\frac{1-q_j+2\pi_j q_j}{1+q_j}$ – without affecting incentives. So $\frac{1-q_{\hat{j}}+2\pi_{\hat{j}}q_{\hat{j}}}{1+q_{\hat{j}}} = \frac{1-q_j+2\pi_j q_j}{1+q_j}$ $\forall j$. This defines the monitoring probability on task j in terms of parameters and the monitoring probability on any other task. For $\pi_j > \pi_k$, $q_j > q_k$: the firm monitors more intensively when the prior is more pronounced, which parallels the finding in Section 5.

The transfer, t, to the worker for acceptable output must satisfy $(1-\left(\frac{1-q_1+2\pi_1 q_1}{1+q_1}\right)^J)\cdot t \geq Jc$. The optimal incentive organization is determined by the system

$$\min_{(q_1,\ldots,q_J)} \left(1+\frac{\sum_{k=1}^J q_k}{J}\right)\cdot\frac{c}{1-\left[\frac{1-q_1+2q_1\pi_1}{1+q_1}\right]^J}$$

subject to

$$q_j = \left(\frac{1-\pi_j}{1-\pi_1}\frac{1+q_1}{q_1}-1\right)^{-1}, \quad \text{for j} = 1,\ldots,\text{J}$$

$$q_j \in (0,1], \quad \text{for j} = 1,\ldots,\text{J}.$$

The objective function states that the firm is minimizing the expected per-task wage bill. The constraints allow the firm to treat this multivariate optimization problem as a univariate problem.

The solution to this problem calls for monitoring the task with the most pronounced prior more intensively than in the optimal incentive organization from Section 5 with $\pi = \max_j \pi_j$. In Section 5, monitoring one task more intensively implied monitoring all tasks more intensively. Here, monitoring the task more intensively allows the firm to decrease the intensity with which it monitors other tasks.

7 Conclusion

New information technology permits firms and workers to interact through spot labor markets. Compared to conventional labor markets, spot markets offer significant advantages for a firm. A flexible and scalable workforce is available to start work immediately and no preexisting relationship with a worker is presumed nor is the promise of a continuing relationship required.

The minimal interaction between the firm and its employees raises new challenges. The firm must provide adequate supervision to ensure workers are acting faithfully on its behalf. Workers are compensated for their effort, but the exertion of effort is costly and unobservable. Furthermore, the quality of a worker's output cannot be verified directly.

With traditional reputation mechanisms inapplicable and the threat of large penalties for poor performance unavailable, the firm creates incentives for effort by periodically hiring additional workers to duplicate some of the tasks it has already assigned. Wages are then made contingent upon satisfactory performance on all tasks.

The firm's monitoring technology is endogenously determined to balance the costs and benefits of supervision. Employing monitors more frequently keeps wages down but leads to greater duplication of assignments. In the optimal organization of incentives, the firm bundles multiple tasks together for each worker. Monitoring and bundling are strategic substitutes: as the firm assigns more tasks to each worker, it monitors each task less frequently.

References

Abreu, D., P. Milgrom, and D. Pearce (1991): "Information and Timing in Repeated Partnerships," *Econometrica*, 59(6), 1713–1733.

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *The Review of Economic Studies*, 58(4), 621–654.

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs and Economic Organization," *American Economic Review*, 62(5), 777–795.

Allen, D. W. (2011): *The Institutional Revolution: Measurement and the Economic Emergence of the Modern World*. University of Chicago Press, Chicago.

Alonso, O., and S. Mizzaro (2009): "Can We Get Rid of TREC Assessors? Using Mechanical Turk for Relevance Assessment," in *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*, ed. by S. Geva, J. Kamps, T. Peters, T. Saka, A. Trotman, and E. Vorhees, pp. 15–16, Amsterdam. IR Publications.

Barrett, W. P. (2011): "Chances Of Tax Audit Small But Growing: Are You A Target?," http://www.forbes.com/sites/williampbarrett/2011/04/01/chances-of-tax-audit-small-but-growing-are-you-a-target/.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169–217.

Becker, G. S., and G. J. Stigler (1974): "Law Enforcement, Malfeasance, and Compensation of Enforcers," *The Journal of Legal Studies*, 3(1), 1+.

Bolton, P., and M. Dewatripont (2005): *Contract Theory*. MIT Press, Cambridge, Massachusetts.

Bose, S., and J. Zhao (2007): "Optimal Use of Correlated Information In Mechanism Design When Full Surplus Extraction May Be Impossible," *Journal of Economic Theory*, 135, 357–381.

Calvo, G. A., and S. Wellisz (1978): "Supervision, Loss of Control, and the Optimum Size of the Firm," *Journal of Political Economy*, 86(5), 943–952.

Carmichael, H. L. (1985): "Is Unemployment Involuntary? Comment," *American Economic Review*, 75, 1213–1214.

Carr-Hill, R. A., and N. H. Stern (1979): *Crime, the Police, and Criminal Statistics*. Academic Press, London: UK.

Caulfield, B. (2011): "Amazon Mechanical Turk Will Hook You Up With Porn Jobs That Pay Pennies," *Forbes*.

CBSNews (2010): "Lowering Your Odds of Facing IRS Audit," http://www.cbsnews.com/2100-500200_162-6311318.html.

Cremer, J., and R. P. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247–1257.

Dickens, W. T., L. F. Katz, K. Lang, and L. H. Summers (1989): "Employee Crime and the Monitoring Puzzle," *Journal of Labor Economics*, 7(3), 331–347.

Farber, D. A. (1980): "Reassessing the Economic Efficiency of Compensatory Damages for Breach of Contract," *Virginia Law Review*, 66, 1443–1484.

Frei, B. (2009): "Paid Crowdsourcing: Current State and Progress Toward Mainstream Business Use," Discussion paper.

Fuchs, W. (2007): "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 97(4), 1432–1448.

Grossman, S. J., R. E. Kihlstrom, and L. J. Mirman (1977): "A Bayesian Approach to the Production of Information and Learning by Doing," *The Review of Economic Studies*, 44(3), 533–547.

Holmstrom, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10(1), 74–91.

——— (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 13(2), 324–340.

Horton, J. J., and L. B. Chilton (2010): "The Labor Economics of Paid Crowdsourcing," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pp. 209–218, New York, NY. ACM.

Horton, J. J., D. G. Rand, and R. Zeckhauser (in press): "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics*.

Investopedia (2010): "Avoid An Audit: 6 "Red Flags" You Should Know," `http://www.investopedia.com/articles/pf/07/avoid_audits.asp`.

Ipeirotis, P. G. (2010): "Demographics of Mechanical Turk," New York University: No. CeDER-10-01.

Kolm, S.-C. (1973): "A Note on Optimum Tax Evasion," *Journal of Public Economics*, 2, 265–270.

Legros, P., and S. A. Matthews (1993): "Efficient and Nearly-Efficient Partnerships," *The Review of Economic Studies*, 60(3), 599–611.

Little, G., L. B. Chilton, M. Goldman, and R. C. Miller (2010): "Exploring Iterative and Parallel Human Computation Processes," in *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 68–76, New York. ACM.

MacLeod, W. B. (2003): "Optimal Contracting with Subjective Evaluation," *American Economic Review*, 93(1), 216–240.

Marge, M., S. Banerjee, and A. I. Rudnicky (2010): "Using the Amazon Mechanical Turk for Transcription of Spoken Language," in *Proceedings of the 2010 IEEE Conference on Acoustics, Speech and Signal Processing*, ed. by J. Hansen, pp. 5270–5273. IEEE.

Mason, W., and S. Suri (2012): "Conducting Behavioral Research on Amazon's Mechanical Turk," *Behav Res Methods*, 44(1), 1–23.

Mason, W. A., and D. J. Watts (2009): "Financial Incentives and the 'Performance of Crowds'," *Knowledge Discovery and Data Mining - Human Computation (KDD-HCOMP)*, 11(2), 100–108.

McAfee, P. R., and P. J. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica*, 60(2), 395–421.

Mirrlees, J. A. (1974): *Essays in Equilibrium Behavior under Uncertainty*chap. Notes on Welfare Economics, Information and Uncertainty. North Holland, Amsterdam.

——— (1975): "The Theory of Moral Hazard and Unobservable Behavior," Mimeo, Oxford.

——— (1976): "The Optimal Structure of Incentives and Authority Within an Organization," *Bell Journal of Economics*, 7, 105–131.

oDesk (2012): "oDesk at a Glance," .

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010): "Running Experiments on Amazon Mechanical Turk," *Judgement and Decision Making*, 5(5).

Posner, R. (1972): *Economic Analysis of Law*. Little, Brown, Boston.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37(1), 7–63.

Rahman, D. (2012): "But Who Will Monitor the Monitor," *American Economic Review*, 102(6), 2767–2797.

Rahman, D., and I. Obara (2010): "Mediated Partnerships," *Econometrica*, 78(1), 285–308.

Riordan, M. H., and D. E. Sappington (1988): "Optimal Contracts with Public Ex Post Information," *Journal of Economic Theory*, 45, 189–199.

Robbins, H. (1952): "Some Aspects of the Sequential Design of Experiments," *Bulletin of the American Mathematical Society*, 55, 527–535.

Ross, J., L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson (2010): "Who are the Crowdworkers? Shifting Demographics in Amazon Mechanical Turk," in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, ed. by K. Edwards, and T. Rodden, pp. 2863–2872, New York. ACM.

Roth, D. L. (1997): *Crazy from the Heat*. Hyperion, New York.

Rothschild, M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9(2), 185–202.

Sanders, D. (2011): "Crowdsourcing Investment," *Crowdsourcing.org*.

Shapiro, C., and J. E. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *The American Economic Review*, 74(3), 433–444.

Snow, R., B. O'Connor, C. Jurafsky, and A. Y. Ng (2008): "Cheap and Fast – but is it good? Evaluating Non-Expert Annotations for Natural Language Tasks," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ed. by M. Lapata, and H. T. Ng, pp. 254–263, New York. ACM.

Stigler, G. J. (1970): "The Optimum Enforcement of Laws," *Journal of Political Economy*, 78, 526–536.

Suri, S., and D. J. Watts (2011): "Cooperation and Contagion in Web-Based, Networked Public Goods Experiments," *PLoS One*, 6(3).

Urbano, J., J. Morato, M. Marrero, and D. Martin (2010): "Crowdsourcing Preference Judgements for Evaluation of Music Similarity Tasks," in *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*, ed. by M. Lease, V. Carvalho, and E. Yilmaz, pp. 9–16, Geneva, Switzerland.

Vanham, P. (2012): "Virtual Working Takes Off in EMs," *Financial Times*.

Williamson, O. E. (1967): "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, 75(2), 123–138.

A Crowdsoucing

This section provides a brief overview of the crowdsourcing labor spot market. Crowdsourcing is the process of delegating work to an undefined group of people (a crowd) through an open call online. Of the dozens of work exchanges where firms can hire workers, Amazon's Mechanical Turk (AMT) is the most prominent. Created in-house in 2005 to find duplicates among the company's product webpages, the service rapidly expanded and by 2007 comprised a pool of more than 100,000 workers in over 100 countries completing various types of tasks, such as transcribing podcasts, rating and tagging images, and writing/rewriting sentences. There are now more than 150,000 jobs available at any time (Caulfield, 2011).

The paid crowdsourcing market has grown considerably since AMT's founding. oDesk, a competing platform, has 2.3 million registered workers and posted half a million jobs in the second quarter of 2012 (oDesk, 2012). Almost 8 million hours of work were performed in that quarter alone and worker earnings on oDesk tallied $250 million in 2011 (Vanham, 2012).

A paid survey conducted on AMT in February 2010 revealed workers from 68 different countries; the United States is most prevalent at 45% followed by India at 34%. Young workers are overrepresented, even when compared to the general population of Internet users. Self-reported education levels are also greater than those of the general populations.

Most workers spend a day or less working on AMT, completing 20-100 jobs and earning $20 or less per week. There exists a high-end of the income distribution with workers earning more than $1,000 per month. More than 20% of Indian workers report AMT as their primary source of income (10% of American workers), with an additional 35% (60% for American workers) using AMT as a secondary source of income. The primary motivation for working on AMT is to earn cash while spending free time fruitfully (60% of American workers and 70% of Indian workers).[22]

Field experiments carried out on AMT suggest workers respond to economic incentives in a predictable fashion (Horton and Chilton, 2010; Horton, Rand, and Zeckhauser, in press; Mason and Watts, 2009; Paolacci, Chandler, and Ipeirotis, 2010). Explicitly informing workers that the accuracy of their responses was being measured and used to determine whether payment for their work would be provided had no discernible effect on either the quality or quantity of output: participants appear to treat their pay as necessarily performance dependent. Several studies have shown that the performance of subject-matter experts can be achieved via crowdsourcing (Alonso and Mizzaro, 2009; Marge, Banerjee, and Rudnicky, 2010; Paolacci, Chandler, and Ipeirotis, 2010; Snow, O'Connor, Jurafsky, and Ng, 2008; Urbano, Morato, Marrero, and Martin, 2010).[23]

Guidelines for creating contracts on AMT are scant and consist of little more than "be clear." The definition of acceptable output and the conditions for payment are not normally presented on the main job posting page. Figure 6 provides an example from an actual job posting on AMT.

Hiring multiple workers is probably the most common method of quality assurance (Ma-

[22]See, also, Mason and Watts (2009) and Suri and Watts (2011) on demographics and Ross, Irani, Silberman, Zaldivar, and Tomlinson (2010) and Ipeirotis (2010) on earnings.

[23]Horton, Rand, and Zeckhauser (in press) and Suri and Watts (2011) recreate laboratory experiments on AMT and find no significant differences between the settings.

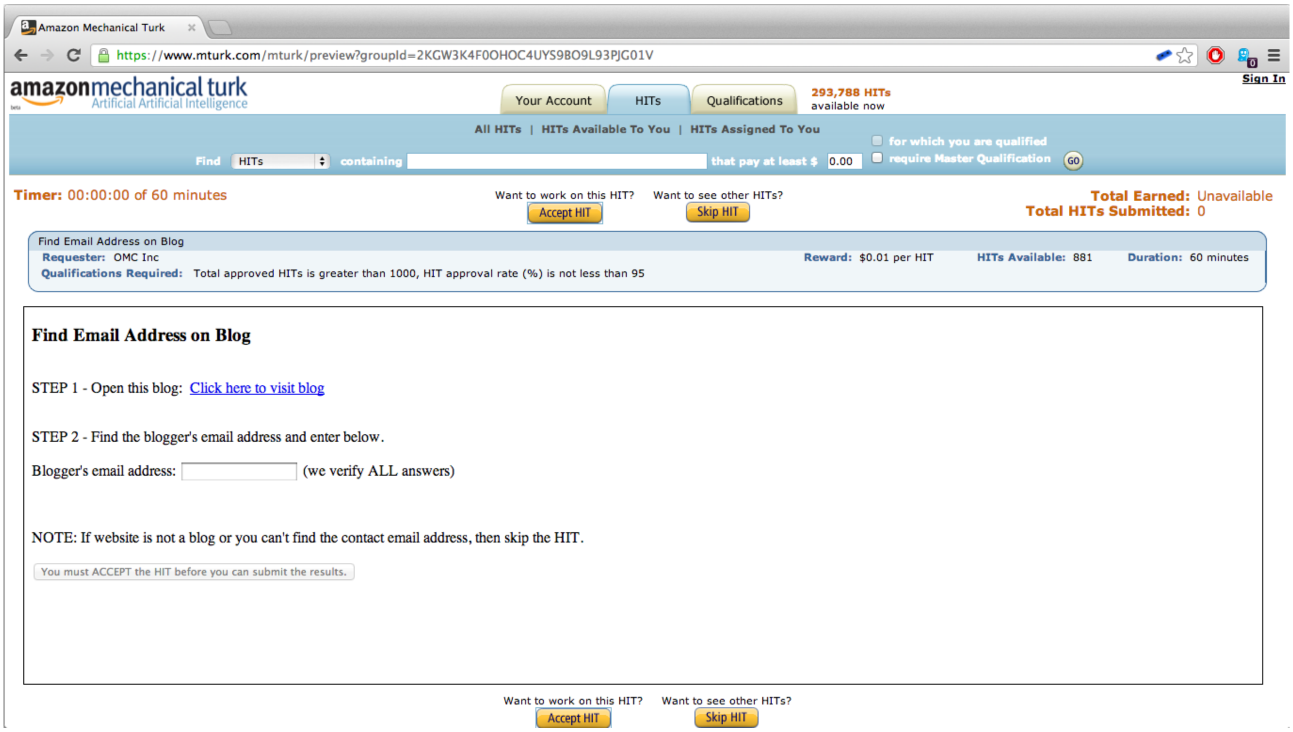

Figure 6: Example job posting on Amazon's Mechanical Turk.

son and Suri, 2012). (The job in Figure 6 states "we [the firm] verify ALL answers.") Obtaining multiple responses is cost-efficient for many tasks on AMT. It also tends to perform well (Snow, O'Connor, Jurafsky, and Ng, 2008). Another option, known as the "Gold Standard," is built around the idea of including tasks within each job for which the firm already knows the correct answer. A worker's performance can then be judged on the subset of these tasks.[24]

[24] Little, Chilton, Goldman, and Miller (2010) find that for tasks with clearly unreasonable answers, firms can effectively vet the responses of workers by employing additional agents for judgement. The additional agents verify whether the first worker provided a coherent response to the question.

B Proofs

Proof of Remark 1
Remark 1. *When effort is contractible, $Q^* = (0, ..., 0)$ and*

$$T^*(r^i) = \begin{cases} J \cdot c & \forall\ r^i\ \in \overline{\mathcal{R}}^i \text{ and } \sigma^{\rm i} = \overline{\sigma}^{\rm i} \\ 0 & \forall\ r^i\ \in \overline{\mathcal{R}}^i \text{ and } \sigma^{\rm i} \neq \overline{\sigma}^{\rm i}. \end{cases}$$

Proof. IR is satisfied with a per-task wage equal to the worker's cost of effort. No monitors need to be employed due to the lack of incentive considerations. □

Proof of Lemma 1
Lemma 1. *For any equilibrium of the firm's problem, there exists another equilibrium providing the same expected payoffs to the firm and all agents in which (a) the firm adopts an implementation plan $Q^i = (q_1, ..., q_J)$ with $q_j = \sum_{n=2}^{\infty} f_j(n)$ and $f_j(n) = 0$ for all $n > 2$, and (b) workers use truthful recommendation strategies $\rho = \overline{\rho}$.*

Proof. In equilibrium, two workers are sufficient to induce high effort on every task. Hiring additional workers on a task changes neither the incentives facing workers nor the probability the firm takes the correct action. The only effect is that the firm's wage bill increases.

The firm correctly interprets recommendations in equilibrium. For any equilibrium with $\rho \neq \overline{\rho}$, there exists another equilibrium where recommendation strategies are $\overline{\rho}$ yielding identical expected payoffs for all actors. □

Proof of Lemma 2
Lemma 2. *The optimal incentive organization sets $T^i(r^i) = 0\ \forall r^i \in \mathcal{R}^i_{kj}$ whenever $k \neq j$.*

Proof. Equation 1 holds

$$t_{00}[\overline{Pr}(\mathcal{R}^i_{00}) - Pr(\mathcal{R}^i_{00})] + t_{10}[\overline{Pr}(\mathcal{R}^i_{10}) - Pr(\mathcal{R}^i_{10})] + ... +$$
$$t_{JJ}[\overline{Pr}(\mathcal{R}^i_{JJ}) - Pr(\mathcal{R}^i_{JJ})]\ \geq\ c[J - \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i)].$$

For any $\mathcal{R}^i_{kj}$ with $k \neq j$, $\overline{Pr}(\mathcal{R}^i_{kj}) = 0$. (If all workers are exerting effort, then a worker necessarily produces matching output whenever monitored.) So $\overline{Pr}(\mathcal{R}^i_{kj}) \leq Pr(\mathcal{R}^i_{kj})$ and setting $T^i(r^i) > 0$ for $r^i \in \mathcal{R}^i_{kj}$ can only make Equation 1 harder to satisfy. □

Proof of Theorem 1

The proof of Theorem 1 is divided into several lemmas. The first lemma shows that a necessary and sufficient condition on the wage is that it discourage strategies $\sigma^i = (p^i_1, ..., p^i_J)$ where $p^i_j \in \{0, 1\}$. The second lemma shows that for any implementation plan $Q^i = (q_1, ..., q_J)$ with some $q_j \neq q_k$, there exists an implementation plan such that $q_j = q$ for $j = 1, ..., J$ providing the same disincentives to shirking at lower expected cost. The final lemma shows that it is sufficient for the firm to discourage $\underline{\sigma}^i$. The wage required to induce effort follows.

Lemma 3. *It is necessary and sufficient for the firm to discourage deterministic effort strategies $\sigma^i = (p_1^i, ..., p_1^J)$ where $p_j^i \in \{0, 1\}$ for all j and i.*

Proof. Necessity is obvious. For sufficiency, the incentive compatibility constraint is given by Equation 3:

$$t \geq \frac{c(J - \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i))}{1 - \left[\frac{1-q_j+2q_j(p_j^i+(1-p_j^i)\pi)}{1+q_j}\right]^J}, \quad \forall \sigma^i \in \Sigma^i.$$

We now show this expression depends linearly on the probability the agent exerts effort on task j, p_j^i. The probability of matching report component j of a coworker is given by $\frac{1-q_j}{1+q_j} + \frac{2q_j}{1+q_j}[p_j^i + (1-p_j^i)\pi]$. Because of independence across tasks, $\left[\frac{1-q_j+2q_j(p_j^i+(1-p_j^i)\pi)}{1+q_j}\right]^J$ may be written as $\Upsilon\big(\frac{1-q_j}{1+q_j} + \frac{2q_j}{1+q_j}[p_j^i + (1-p_j^i)\pi]\big)$ for some Υ. Likewise, the expected total cost of effort, $\sum_{k=0}^{J} k \cdot c \cdot \zeta(k|\sigma^i)$, can be written as $\sum_{k=0}^{J-1} k \cdot c \cdot \zeta(k|\sigma^i) + p_j^i \cdot c = \Xi \cdot c + p_j^i \cdot c$ for some Ξ. The incentive constraint must then satisfy

$$t \geq \frac{c \cdot (J - \Xi - p_j^i)}{1 - \Upsilon\big(\frac{1-q_j}{1+q_j} + \frac{2q_j}{1+q_j}[p_j^i + (1-p_j^i)\pi]\big)}.$$

Consider how this transfer changes as p_j^i varies. The worker can select any $p_j^i \in [0, 1]$, so the payment must dissuade all such effort choices. The direction of the change in the required transfer as p_j^i varies, $\frac{\partial t^i}{\partial p_j^i}$, is given by the sign of

$$-\left(1 - \Upsilon\left(\frac{1-q_j}{1+q_j} + \frac{2q_j}{1+q_j}[p_j^i + (1-p_j^i)\pi]\right)\right) + V\frac{2q_j}{1+q_j}(1-\pi)(J - \Xi - p_j^i),$$

which changes linearly in p_j^i. The required wage is largest either when $p_j^i = 0$ or $p_j^i = 1$. This shows that it is necessary and sufficient for the firm to offer a large enough transfer to dissuade the worker from shirking on $n > 0$ of the J tasks assigned; it is without loss of generality to restrict the worker to deterministic strategies $\sigma^i = (p_1^i, ..., p_J^i)$ with $p_j^i \in \{0, 1\}$ for all j. □

Lemma 4. *For any implementation plan $Q = (q_1, ..., q_J)$ with $q_j \neq q_k$ for some j, k, there exists another implementation plan $\hat{Q} = (q, ..., q)$ providing the same disincentives to shirking at lower expected cost.*

Proof. From Lemma 3, the incentive constraint requires

$$\overline{Pr}(\overline{\mathcal{R}}^i) \cdot t - Jc \geq Pr(\overline{\mathcal{R}}^i) \cdot t - nc$$

for all strategies σ^i that specify shirking on exactly $n = 0, ..., J$ tasks. Since $\overline{Pr}(\overline{\mathcal{R}}^i) = 1$, the required transfer is governed by $Pr(\overline{\mathcal{R}}^i)$, which itself depends on the firm's implementation plan.

Let $Q = (q_1, ..., q_J)$ be given and consider strategy σ^i in which i shirks on n tasks and the firm's monitoring probability on each such task is q_k, $k = 1, ..., n$. Then $Pr(\overline{\mathcal{R}}^i) = \left[\frac{1-q_1+2q_1\pi}{1+q_1}\right] \cdot ... \cdot \left[\frac{1-q_n+2q_n\pi}{1+q_n}\right]$.

For any collection of monitoring probabilities $(q_1, ..., q_J)$, the firm can set a uniform monitoring probability q such that $Pr(\overline{\mathcal{R}}^i)$ is the same. In other words, there exists q such that

$$\left[\frac{1-q+2q\pi}{1+q}\right]^n = \left[\frac{1-q_1+2q_1\pi}{1+q_1}\right] \cdot ... \cdot \left[\frac{1-q_n+2q_n\pi}{1+q_n}\right].$$

Formally, $\left[\frac{1-q+2q\pi}{1+q}\right]$ is the geometric mean of $\left[\frac{1-q_1+2q_1\pi}{1+q_1}\right] \cdot ... \cdot \left[\frac{1-q_n+2q_n\pi}{1+q_n}\right]$.

The firm cares about the expected wage bill. For each task j, $1 + q_j$ workers are being paid in expectation, so the firm expects to pay $J + \sum_{k=1}^J q_k$ workers in total. Suppose under implementation plan $Q = (q_1, ..., q_J)$, $q_1 > q_2$, so that $\frac{1-q_1+2q_1\pi}{1+q_1} < \frac{1-q_2+2q_2\pi}{1+q_2}$ and $\ln\left(\frac{1-q_1+2q_1\pi}{1+q_1}\right) < \ln\left(\frac{1-q_2+2q_2\pi}{1+q_2}\right)$.

Consider a geometric-mean-preserving scrunch of $\left[\frac{1-q_1+2q_1\pi}{1+q_1}\right] \cdot ... \cdot \left[\frac{1-q_n+2q_n\pi}{1+q_n}\right]$ obtained by decreasing q_1 and simultaneously increasing q_2. By definition, the scrunch preserves the average of the logs $\frac{1}{n}\sum_{j=1}^n \ln\left(\frac{1-q_j+2q_j\pi}{1+q_j}\right)$ as well as $Pr(\overline{\mathcal{R}}^i)$.

Since $\ln\left(\frac{1-q_j+2q_j\pi}{1+q_j}\right)$ is decreasing and concave in q_j, the geometric-mean-preserving scrunch in which q_1 is decreased and q_2 is increased permits q_1 to fall by more than q_2 rises. Thus, the sum $\sum_{k=1}^J q_k$ decreases. The implementation plan resulting from the scrunch provides the same transfer and discourages shirking exactly as effectively as before. But since it has reduced the expected number of workers hired, it results in a lower expected wage bill. The firm's optimal implementation plan must then be of the form $\hat{Q} = (q, ..., q)$ specifying a constant monitoring probability $q \in [0, 1]$. □

Lemma 5. *It is necessary and sufficient that the transfer dissuades shirking on all tasks. In other words, the incentive compatible wage is determined by* $\underline{\sigma}^i$.

Proof. Let σ^i call for shirking on exactly $J - n$ tasks (equivalent to exerting effort on n tasks). Implementation plan $Q = (q, ..., q)$ implies $Pr(\overline{\mathcal{R}}^i) = \left[\frac{1-q+2q\pi}{1+q}\right]^{J-n}$. The incentive compatibility constraint can now be written as

$$t \geq \frac{c(J-n)}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^{J-n}}.$$

Define $\nu = \frac{1-q+2q\pi}{1+q}$ and $x = J - n$, the constraint can be rewritten as $t \geq c\frac{x}{1-\nu^x}$. The right-hand side of this expression is increasing in x since the sign of $\frac{\partial}{\partial x}\left(\frac{x}{1-\nu^x}\right)$ is given by $1 - \nu^x + \nu^x \ln(\nu^x)$, which is positive. (The expression $1 - \nu^x + \nu^x \ln(\nu^x)$ is decreasing in ν, but it is still positive as $\nu^x \to 1$ from below.)

The transfer must be large enough to dissuade shirking on all tasks, so

$$t \geq \frac{Jc}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J}.$$

□

Theorem 1. *The optimal contract will take the form of an implementation plan $Q^i = (q, ..., q)$ that specifies hiring a second agent with probability q for each task and a transfer*

$$T^i(r^i) = \begin{cases} J \cdot w^*(q, \pi, c, J) & \forall\ r^i \in \overline{\mathcal{R}}^i \\ 0 & \forall\ r^i \notin \overline{\mathcal{R}}^i, \end{cases}$$

where the equivalent per-task wage is $w^*(q, \pi, c, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$.

Proof. The firm will select the lowest transfer capable of inducing effort on all tasks. Lemma 3 through Lemma 5 establish properties of the optimal contract. For any implementation plan $Q^i = (q, ..., q)$, the transfer scheme specifies $T^i(r^i) = \frac{Jc}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$ for reports $r^i \in \overline{\mathcal{R}}^i$ and $T^i(r^i) = 0$ for $r^i \notin \overline{\mathcal{R}}^i$. The lump-sum transfer is equivalent to a per-task wage of $w(q, \pi, c, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$. □

Proof of Corollary 1

Corollary 1. *$w^*(q, \pi, c, J)$ is decreasing in q and J and increasing in π and c.*

Proof. The claims follow immediately from consideration of $w^*(q, \pi, c, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$. □

Proof of Corollary 2

Corollary 2. *For effort costs $c \in ((1-\pi)\frac{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}{1+q}, 1-\pi]$ the firm will not employ any workers despite it being efficient to do so were shirking not a concern.*

Proof. The firm is willing to pay the equivalent per-task wage $w^*(q, \pi, c, J)$ instead of hiring no worker only if

$$1 - (1+q)\frac{c}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J} \geq \pi.$$

The left-hand side gives the firm's expected per-task payoffs from the proposed contract while the right-hand side is the firm's expected per-task payoffs from guessing the state. This condition is equivalent to

$$c \leq (1-\pi)\frac{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J}{1+q} < 1 - \pi.$$

The relevant condition when effort is contractible is $c \leq 1 - \pi$. Thus, for effort costs $c \in ((1-\pi)\frac{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}{1+q}, 1-\pi]$ the firm will not employ any workers despite hiring workers when effort is contractible. □

Proof of Theorem 2
Theorem 2. $q^*(\pi, c, J) < 1$ *if and only if* $J > 1$.

Proof. The firm's expected payoff per task is

$$1 - (1+q)\left(\frac{c}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J}\right).$$

Let (π, c, J) be given and maximize this payoff with respect to q. To simplify notation, let $\nu = \frac{1-q+2q\pi}{1+q}$ be the probability of producing acceptable output on any task when shirking. Note that $\nu \in [\pi, 1]$ is decreasing over $q = 0$ to $q = 1$. The first-order condition stipulates

$$c \cdot \frac{\left[-2J(\pi-1)\nu^J + \nu^J + (2\pi-1)q(\nu^J - 1) - 1\right]}{[(2\pi-1)q+1][\nu^J - 1]^2} = 0,$$

which requires

$$-2J(\pi-1)\nu^J + \nu^J + (2\pi-1)q(\nu^J-1) - 1 = 0. \tag{5}$$

When $J = 1$, $\nu = \pi$ satisfies Equation 5, so $q^*(\pi, c, 1) = 1$.

Suppose $J > 1$. When $q = 0$, the left-hand side of Equation 5 is positive. It is negative when $q = 1$. To see this, note that the left-hand side reduces to $-2J\pi^{J+1} + 2J\pi^J + 2\pi^{J+1} - 2\pi$ at $q = 1$. This expression is less than zero for $-J\pi^J + J\pi^{J-1} + \pi^J < 1$. The sum $-J\pi^J + J\pi^{J-1} + \pi^J$ is strictly increasing in π and equals unity at $\pi = 1$, so it is less than unity for $\pi < 1$. Thus, the left-hand side of Equation 5 is negative when $q = 1$ for any $\pi < 1$. Since Equation 5 is continuous in q, it is satisfied with equality at some interior q^* by application of the intermediate value theorem. The second-order condition assures q^* identifies the unique maximum of the firm's optimization problem. □

Proof of Theorem 3
Theorem 3. $q^*(\pi, c, J) \to 0$ *as* $J \to \infty$.

Proof. Let (π, c) be given. The dependence of the monitoring probability on these parameters will often be suppressed. Denote $\nu = \frac{1-q+2q\pi}{1+q}$. Rearranging Equation 5 in the proof of Theorem 2, the optimal monitoring probability q must satisfy

$$\frac{J\nu^J}{1-\nu^J} - \frac{1-q+2q\pi}{2(1-\pi)} = 0. \tag{6}$$

Denote the left-hand side of Equation 6 as $f(J, q(J))$ and treat J as a continuous variable. (The firm's problem is well-defined for $J \in \mathbb{R}_+$ and all relevant objects are monotonic in J.) The partial derivative of f with respect to J is $\frac{\nu^J(-\nu^J + J\ln(\nu)+1)}{(\nu^J-1)^2}$. The sign of this derivative is governed by $1 - \nu^J + J\ln(\nu) \equiv 1 - x + \ln(x)$ where $x = \nu^J$. The sum $1 - x + \ln(x)$ is increasing over $x \in (0, 1]$, so it is increasing in ν^J. Since $\lim_{x\to 1} 1 - x + \ln(x) = 0$, $1 - \nu^J + \ln(\nu^J)$ is negative. Thus, $f_J < 0$. The partial derivative of f with respect to q, f_q, is also negative. Implicit differentiation holds $q_J = \frac{-f_J}{f_q}$, which shows the optimal monitoring probability is decreasing in the job size.

Since $q^*(J) \equiv q^*(\pi, c, J)$ is decreasing in J and bounded below, $\hat{q} = \lim_{J\to\infty} q^*(J)$ exists. To see that the limiting monitoring probability is zero, suppose otherwise that $\hat{q} > 0$. Define $\nu^*(J) = \frac{1-q^*(J)+2\pi q^*(J)}{1+q^*(J)}$. Since $\hat{q}$ exists, $\hat{\nu} = \lim_{J\to\infty} \nu^*(J)$ exists.

Equation 5 implies $-2J(\pi-1)\nu^*(J)^J + \nu^*(J)^J + (2\pi-1)q^*(J)(\nu^*(J)^J - 1) = 1$ must hold at each J. This requires $\lim_{J\to\infty} J\nu^*(J)^J$ to converge to a finite value, implying $\lim_{J\to\infty} \nu^*(J)^J = 0$.

Then, for J sufficiently large, $2J(1-\pi)\nu^*(J)^J - (2\pi-1)q^*(J) \approx 1$ and so $\lim_{J\to\infty} J\nu^*(J)^J = \lim_{J\to\infty} \frac{1-q^*(J)+2\pi q^*(J)}{2(1-\pi)} = \lim_{J\to\infty} \frac{1+q^*(J)}{2(1-\pi)}\nu^*(J)$ and $\lim_{J\to\infty} J\nu^*(J)^{J-1} = \lim_{J\to\infty} \frac{1+q^*(J)}{2(1-\pi)}$.

Let $\varepsilon > 0$ be given. Since $\nu^*(J)$ is convergent, for J large $(J-1)\nu^*(J-1)^{J-1} > (J-1)\nu^*(J)^{J-1} - \frac{\varepsilon}{2}$. Likewise, since $J\nu^*(J)^J$ is convergent, it is Cauchy convergent and there exists J sufficiently large so that $(J-1)\nu^*(J-1)^{J-1} - J\nu^*(J)^J < \frac{\varepsilon}{2}$. Thus

$$
\begin{aligned}
\frac{\varepsilon}{2} &> (J-1)\nu^*(J-1)^{J-1} - J\nu^*(J)^J \\
&> (J-1)\nu^*(J)^{J-1} - \frac{\varepsilon}{2} - J\nu^*(J)^J \\
&= \frac{1+q^*(J)}{2(1-\pi)} - \nu^*(J)^{J-1} - J\nu^*(J)^J - \frac{\varepsilon}{2} \\
&= \frac{1+q^*(J)}{2(1-\pi)} - \nu^*(J)^{J-1} - \frac{1+q^*(J)}{2(1-\pi)}\nu^*(J) - \frac{\varepsilon}{2} \\
&= \frac{1+q^*(J)}{2(1-\pi)}[1-\nu^*(J)] - \nu^*(J)^{J-1} - \frac{\varepsilon}{2}.
\end{aligned}
$$

So for J sufficiently large, $\varepsilon > \frac{1+q^*(J)}{2(1-\pi)}[1-\nu^*(J)] - \nu^*(J)^{J-1}$. But since $\nu^*(J)^J \to 0$, this means $1 = \lim_{J\to\infty} \nu^*(J) = \lim_{J\to\infty} \frac{1-q^*(J)+2\pi q^*(J)}{1+q^*(J)}$. Recalling that $\pi < 1$ was given, this requires $q^*(J) \to 0$. This contradicts the maintained hypothesis that $\hat{q} > 0$ and the limiting monitoring probability is zero.

Therefore $q^*(\pi, c, J) \to 0$ as $J \to \infty$. □

Proof of Corollary 3

Corollary 3. *For any (π, c), given $\varepsilon > 0$, $\exists\ J_\varepsilon < \infty$ such that $q^*(\pi, c, J) > \varepsilon$ for all $J < J_\varepsilon$.*

Proof. The result is immediate from noting the firm's per-task wage bill as $q \to 0$ holding fixed J grows without bound: $\lim_{q\to 0}(1+q) \cdot \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$ does not exist. □

Proof of Result 1

Result 1. *As $J \to \infty$, $(1+q^*(\pi, c, J)) \cdot w^*(q^*(\pi, c, J), \pi, c, J) \to c$.*

Proof. The proof of Theorem 3 shows $\left[\frac{1-q^*(\pi,c,J)+2\pi q^*(\pi,c,J)}{1+q^*(\pi,c,J)}\right]^J \to 0$ as $J \to \infty$. Thus, $w^*(q^*(\pi, c, J), \pi, c, J) \to c$ as $J \to \infty$. Since $q^*(\pi, c, J) \to 0$ as $J \to \infty$, $(1+q^*(\pi, c, J)) \cdot w^*(q^*(\pi, c, J), \pi, c, J) \to c$ as $J \to \infty$. □

Result 3. *The transfer scheme in the firm's optimal contract without B is*

$$T^i(r^i) = \begin{cases} J \cdot w^*(q, \pi, c, J) & \forall\ r^i\ \in \mathcal{R}^i_{JJ} \\ 0 & \forall\ r^i\ \notin \mathcal{R}^i_{JJ}, \end{cases}$$

where the equivalent per-task wage is $w^*(q, \pi, c, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$.

Proof. Lemma 2 continues to apply for the same reasons as before. The general incentive constraint remains

$$t_{00}\big[\overline{Pr}(\mathcal{R}^i_{00}) - Pr(\mathcal{R}^i_{00})\big] + ... + t_{JJ}\big[\overline{Pr}(\mathcal{R}^i_{JJ}) - Pr(\mathcal{R}^i_{JJ})\big] \geq c\big[J - \sum_{k=0}^{J} k \cdot \zeta(k|\sigma^i)\big].$$

Since $\overline{Pr}(\mathcal{R}^i_{jj}) - Pr(\mathcal{R}^i_{jj})$ is largest at $j = J$, inserting a wedge between $\overline{\sigma}^i$ and σ^i is most effective here. A wage of t_{JJ} payable only for reports $r^i \in \mathcal{R}^i_{JJ}$ is capable of inducing effort, where $t_{JJ} = \frac{c\left[J-\sum_{k=0}^{J} k\cdot\zeta(k|\sigma^i)\right]}{\overline{Pr}(\mathcal{R}^i_{jj})-Pr(\mathcal{R}^i_{jj})}$. Paying transfer t for other reports allows t_{JJ} to be lowered, but the expected wage bill increases nonetheless. □