Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

***PIER Working Paper 16-023***

# Optimal Contracting with Costly State Verification, with an Application to Crowdsourcing

by

J. Aislinn Bohren

Troy Kravitz

http://ssrn.com/abstract=2889031

# Optimal Contracting with Costly State Verification, with an Application to Crowdsourcing*


J. Aislinn Bohren† University of Pennsylvania

Troy Kravitz‡ Federal Deposit Insurance Corp.


June 15, 2016


**Abstract**

A firm employs workers to obtain costly unverifiable information – for example, categorizing the content of images. Workers are monitored by comparing their messages. The optimal contract under limited liability exhibits three key features: (i) the monitoring technology depends crucially on the commitment power of the firm – virtual monitoring, or monitoring with arbitrarily small probability, is optimal when the firm can commit to truthfully reveal messages from other workers, while monitoring with strictly positive probability is optimal when the firm can hide messages (partial commitment), (ii) bundling multiple tasks reduces worker rents and monitoring inefficiencies; and (iii) the optimal contract is approximately efficient under full but not partial commitment. We conclude with an application to crowdsourcing platforms, and characterize the optimal contract for tasks found on these platforms.



*We thank Nageeb Ali, Cécile Aubert, Gordon Dahl, Roger Gordon, Kevin Lingerfelt, Mark Machina, George Mailath, Steve Mattthews, David Miller, David Rahman, Branislav Slantchev, Joel Sobel, Ross Starr, Joel Watson and seminar participants at the University of Chicago, Johns Hopkins University, Stanford University, UCSD, U. Pennsylvania and Yale.



†Email: abohren@sas.upenn.edu; Web: https://sites.google.com/site/aislinnbohren/

‡Email: tkravitz@fdic.gov

# 1 Introduction

Technological advancement is changing how modern companies interact with their workforce. New information technology, such as crowdsourcing platforms, creates the opportunity for firms to access a flexible and inexpensive pool of temporary workers on-demand. Millions of potential employees are available around the clock and may start work immediately. Crowdsourcing is already big business: worker earnings are in the billions and the revenues of platforms matching firms to workers were estimated at $500 million in 2009 (Frei 2009). The allure of these online labor markets are their low frictions. But in the absence of conventional methods of supervising employees, tapping into this online workforce presents a host of new incentive issues. The firm must guard against shirking, but how?

We characterize how firms should optimally structure incentives when hiring in a spot market. To fix ideas, suppose a firm has a series of images, and each image is either a tree or a flower. The firm hires workers to accurately categorize these images. Importantly, in this setting, the effort cost of checking a task is similar to the effort cost of completing the task. For example, verifying that an image is correctly categorized requires the same effort as categorizing the image.[1] If the worker exerts costly effort to view the image, he observes the correct category. The firm does not observe whether a worker views the image and the worker does not directly care about whether his categorization is correct. After choosing whether to view the image and observing the category, the worker sends a message stating whether the image is a tree or a flower. The information a worker observes is not verifiable, but the firm can assess the accuracy of a worker's message by assigning the same image to other workers and comparing the messages. Incentives are generated by conditioning a worker's payment upon this comparison, which we refer to as *peer-monitoring*.[2] A key feature of this environment is that the cost of monitoring is endogenous, since the incentive compatible wage payments depend crucially on the probability a second worker

[1] In contrast, verifying the accuracy of a computer program can be significantly less involved than writing the program in the first place.

[2] We discuss other types of monitoring, such as inserting known images into the pool of tasks, in the crowdsourcing application, and demonstrate that peer-monitoring is outperforms these commonly used monitoring schemes under a broad set of conditions.

is assigned to each image.

In addition to ensuring a contract effectively monitors workers, the firm must also ensure that the contract is credible. First, it seems unlikely that the firm would be able to assess penalties for poor performance when workers can simply disappear from the platform. Therefore, the most severe punishment is withholding wages, and workers are protected by limited liability. Second, the firm conditions a worker's payment on the messages of other workers, which a worker does not directly observe. Therefore, whether the firm can commit to truthfully reveal the messages of other workers will affect the credibility of a contract. For example, if its possible to hide the messages of other workers and pay the worker a lower wage, the firm will do so. We focus on a simple departure from full commitment in which the firm can hide messages from a worker's peers but cannot fabricate the content of these messages (i.e. the messages a firm observes are verifiable). We compare this partial commitment setting to one with full commitment.

Our main result is to derive the firm's optimal contract. The optimal monitoring rate and wage scheme depend crucially on the commitment power of the firm. Three key features emerge: (i) virtual monitoring, or monitoring with arbitrarily small probability, is optimal under full commitment while stochastic monitoring, or monitoring with strictly positive probability, is optimal under partial commitment, (ii) bundling – simultaneously assigning a worker multiple images – reduces worker rents and the inefficiency that arises from monitoring, and (iii) the optimal contract is approximately efficienct under full commitment but not under partial commitment.

The firm's commitment power determines what type of wage schemes are credible. When the firm has full commitment, it can credibly commit to condition payment on whether or not the worker is monitored. In this case, the firm will pay a large positive wage when a worker is monitored and matches on every assigned image, including an additional bonus for less likely image categories, and otherwise pays a wage of zero. This wage is inversely proportional to the probability of being paid, and therefore a worker's expected wage payment is independent of the monitoring rate.

In contrast, under partial commitment, the former wage scheme is not credible as the firm would not truthfully reveal when a worker is monitored. Instead, the firm pays a positive wage when the worker matches on all images that are monitored, including an additional bonus for less likely image categories. If the worker has any mismatches, he is paid nothing. Compared to full commitment, the worker is paid a lower wage with higher probability. Monitoring at a higher rate lowers the incentive compatible wage for each message but does not affect the probability of payment. Therefore, the expected wage payment to a worker is decreasing in the monitoring rate.[3]

The firm chooses the optimal monitoring rate to minimize the expected wage payment per task, which is the product of the expected number of workers hired and the expected wage payment per worker. Monitoring more frequently increases the former. Under full commitment, the latter is independent of the monitoring rate. Therefore, *virtual monitoring*, or monitoring with arbitrarily low probability, is optimal. This contract is *approximately efficient*, since duplication occurs with arbitrarily small probability (in the efficient contract, a single worker exerts effort on each task).

Under partial commitment, the firm faces a trade-off between efficiency and worker rents. Monitoring at a higher rate reduces the expected payment per worker but increases the expected number of workers hired. *Stochastic monitoring*, or monitoring with strictly positive probability, is optimal. The firm incurs some inefficiency in order to reduce the rents paid to workers. Therefore, the optimal contract is *inefficient*, despite the fact that it is possible to implement the efficient action profile.[4]

*Bundling*, or grouping multiple images together for a single worker, is optimal under both partial and full commitment. Bundling ties the worker's wage for one image to his performance on all images, which partially relaxes the limited liability constraint by allowing the firm to punish a worker on multiple images when he

[3]In contrast, under full commitment, monitoring at a higher rate lowers the incentive compatible wage but is exactly offset by the higher probability of payment, rendering the expected wage payment independent of the monitoring rate.

[4]Note that while information aggregation is a commonly cited motivation for hiring multiple agents, in this model, there is no learning justification for the duplication of tasks: multiple workers are employed with strictly positive probability solely for incentive reasons. Thus, the paper provides a complementary explanation for why a principal may consult multiple agents before taking a decision.

deviates on one. This reduces the rents captured by the workers. Under partial commitment, bundling also reduces the optimal monitoring rate, which reduces the inefficiency that arises from monitoring. Monitoring and bundling are strategic substitutes: the firm hires monitors less frequently as the number of images assigned to an agent increases. The efficiency loss relative to the contractible effort benchmark vanishes asymptotically.

We conclude with an application to an online crowdsourcing marketplace in which the firm offers a simple contract that either accepts or rejects a worker's output. The optimal contract we characterize offers potentially significant improvements over those currently used in practice. Firms can reduce their monitoring expenditures by structuring contracts so that individual workers effectively monitor each other. By shifting from piece-rate payment schedules to schemes requiring satisfactory performance on all tasks, firms can recreate the same incentives at a lower cost. We show that symmetric peer-to-peer monitoring outperforms other types of monitoring, such as inserting known images into the pool (referred to as the Gold Standard in crowdsourcing markets) or constructing a hierarchy of monitors by designating some workers as monitors whose sole responsibility is to check the work of others. While Alchian and Demsetz (1972) suggest that there should be specialization in monitoring, this is not the case in our setting: in the optimal contract, the firm treats workers symmetrically. Importantly, the theoretical improvements we identify can be tested empirically through field experiments on crowdsourcing platforms.

Our main findings apply to settings with multidimensional tasks as well. For example, consider a multidimensional chore in which the firm monitors individual components of the chore and punishes workers across all dimensions for poor performance on any dimension. Bundling these different tasks strengthens incentives relative to contracting each dimension separately. Further applications are discussed in Section 6

**Literature.** In the standard agency model, an agent's action is not observable and he is subject to moral hazard. The firm observes an informative signal about an agent's action and conditions payment on the realization of the signal (Holmstrom 1979; Mirrlees 1976, 1999). As long as the distribution of the signal varies with the chosen action, the firm is able to align incentives so that the agent chooses the desired action.

A complication arises if the firm is restricted in how severely it can punish an employee. Under limited liability, 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, but the worker captures rents (Bolton and Dewatripont 2005, Ch. 4). Workers in our model capture rents for the same reason. The size of these rents is endogenously determined by the firm's choice of monitoring technology.

In multilateral contracting environments, issues arise with statistically distinguishing deviations by individual agents. If the signal only reveals aggregate information about the actions of the group, and not information about individual action choices, the principal must guard against a free-rider problem. Holmstrom (1982) emphasizes the role of group penalties: all workers are punished whenever bad signals obtain. Group penalties are natural in our model since identifying the deviator requires hiring additional workers, which is more costly than identifying that a deviation occurred.

The signal structure can be viewed as the firm's monitoring technology. In much of the literature, the monitoring technology is exogenous. In contrast, the firm chooses the monitoring technology in our paper. Early studies in which monitoring is a choice variable include Becker (1968), Kolm (1973) and Mirrlees (1974).[5] These papers suggest combining infinitesimal monitoring with arbitrarily harsh punishments. Legros and Matthews (1993) study a multilateral partnership problem in which each agent privately devotes effort to a common project. Workers are monitored by instructing one worker to choose an inefficient action with small probability $p$. The cost of monitoring is determined by the loss from playing this inefficient action, and depends on $p$. Under unlimited liability, efficiency is approximated by choosing $p$ close to zero. These requires large fines for some realizations of output. Under limited liability, the smallest $p$ that satisfies incentives is bounded away from zero and the optimal contract is

[5] 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).

inefficient. In our paper, whether limited liability creates an inefficiency depends on the firm's commitment power: the optimal contract under full commitment is approximately efficient but the optimal contract under partial commitment is not.

Rahman (2012) also considers a costly monitoring setting. The first-best strategy profile is for a worker to exert effort and the firm to never monitor. He examines when virtual monitoring is feasible, and shows that the first-best strategy profile can be approximated arbitrarily well by monitoring with low frequency and punishing the worker severely whenever the monitor reports that he shirked. In our setting, virtual monitoring is always feasible. The firm can approximate the first-best action profile, provided there is no upper bound on transfers. However, this is achieved at great cost: wage payments are unboundedly large. With full commitment, the probability of paying these large wage payments shrinks proportionally, so that the expected wage payment is independent of the monitoring probability. In contrast, with partial commitment, the probability of paying a positive wage for a message is independent of the monitoring probability, while the required wage payment becomes unboundedly large as the monitoring probability shrinks. Virtual monitoring is optimal under full commitment but not under partial commitment.

In our paper, bundling multiple tasks strengthens incentives 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. He shows it is optimal for the firm to withhold payment until the final period.[6]

The paper proceeds as follows. An example is presented in Section 2. The formal model is introduced in Section 3, while Section 4 derives the firm's optimal contract under partial and full commitment. Section 5 develops an application to crowdsourcing platforms. Omitted proofs are in the Appendix.

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

## 2 Example

Users upload thousands of images to a social media site each day. The site needs to ensure that the images meet certain guidelines, and hires workers on a large online labor platform to moderate content. Workers receive images and report their content. Viewing the image takes time, which costs 0.02 per image, and accurately reveals objectionable content. Multiple images can be packaged and sent to a worker as a single job. The moderator limits each job to a maximum of 10 images, to ensure approval of listings within a reasonable time frame. Workers and the site know that most images are acceptable: 95% meet guidelines. The risk-neutral site 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 payment to workers. Risk-neutral workers do not care whether the site treats an image correctly: payoffs are wages less the cost of effort.

If the site could observe whether a worker views an image, it would delegate the task and pay the worker his cost of effort, 0.02. But it is not possible to observe whether a worker views an image. The site must design a monitoring technology to prevent workers from shirking and fabricating messages. We consider peer-monitoring, in which the site probabilistically assigns a task to two workers and compares their answers. Denote the probability that an image in a worker's job is assigned to a second worker as $q$. We identify how the site optimally structures incentives under several different contracting environments. A contract consists of the number of images in a job, a payment scheme and a monitoring technology.

**Benchmark: Unlimited Liability.** Suppose the site has no restrictions on negative payments. Then the optimal transfer is to pay a worker 0.02 unless a second worker is hired *and* their messages mismatch, in which case the worker pays a penalty of $0.02(1 - 1/.05q)$. The expected number of workers hired is increasing in $q$, while the equilibrium transfer per worker is independent of $q$. Therefore *virtual monitoring*, or monitoring with arbitrarily small probability, is optimal. The optimal contract is approximately efficient and workers earn no rents. There is no benefit to assigning multiple images to a single worker.

**Limited Liability and Full Commitment.** Suppose that the moderator cannot threaten workers with negative wage payments. In any contract that offers a positive wage for at least one possible message, the worker can earn a positive expected payoff by shirking and fabricating a message. Thus, if the contract provides incentives for the worker to view the image, the worker will capture a premium above his cost of effort. This premium is a general feature of limited liability with unobservable effort.

Suppose a worker is assigned a single image, and this image is assigned to a second worker with probability $q > 0$. In the optimal payment scheme, a worker is paid if and only if a second agent is hired and their messages match; otherwise, the worker is paid nothing. The worker is paid $w_0 = 0.2/0.05q \approx .4/q$ for being monitored and matching when he flags an image and $w_1 = 0.2/0.95q \approx .21/q$ for being monitored and matching when he does not flag an image. The worker bears the risk of monitoring, in the sense that he is paid only when he is monitored. This is a new feature under limited liability; with unlimited liability, there is no benefit to conditioning a worker's payment on the realization of monitoring.

Consider the optimal monitoring probability. With probability $.05q$ the worker meets the criteria for payment and receives wage $w_0$, while with probability $.95q$ the worker meets the criteria for payment and receives wage $w_1$. This yields an expected wage payment per worker of $.02 \times 2 = .04$. Although a lower monitoring rate increases the wage conditional on payment, it also reduces the probability of payment. These two effects exactly offset each other, yielding an expected wage payment per worker that is independent of $q$. Therefore, the optimal monitoring probability minimizes the expected number of workers hired. Once again, virtual monitoring is optimal, and the optimal contract is approximately efficient. The total expected payment approaches $(1 + \varepsilon) \times 0.04$, so workers capture significant rents.

The moderator can improve upon the single image contract by assigning multiple images to workers. Suppose a worker is assigned 10 images and each image is independently assigned to a second worker with probability $q > 0$. A worker is paid if and only if a second agent is hired and their messages match on every assigned task, otherwise, the worker is paid nothing. Virtual monitoring remains optimal, and the moderator's expected payments per image approach

$0.02 \times 2^{10}/(2^{10} - 1) \approx 0.02$ (see Theorem 1). *Bundling*, or assigning a worker multiple images, significantly reduces the rents captured by workers. By assigning multiple images to each worker, the moderator is able to tie a worker's compensation on one task to successful completion of all tasks. This relaxes the limited liability constraint.

**Limited Liability and Partial Commitment.** The previous contract relies on the firm's ability to truthfully disclose messages from other workers. This is often not a realistic assumption, particularly when workers are from a large, online labor market. As a first step in relaxing the commitment power of the firm, suppose that messages are verifiable but the firm cannot commit to revealing the messages of other workers. In other words, the firm can withhold messages but it cannot fabricate messages. We refer to this as partial commitment.

Partial commitment means that a worker must be paid a weakly higher wage for any report that can be generated by hiding some or all of the messages of other workers. The wage structure in the full commitment environment does not satisfy this constraint, as the firm only pays the worker when she is monitored on all tasks. Under partial commitment, the firm could hide the fact that it monitored the worker and pay the worker nothing.

Suppose a worker is assigned a single image, and this image is assigned to a second worker with probability $q > 0$. In the optimal payment scheme, a worker is paid if he is not monitored or is monitored and matches.

Consider the optimal monitoring probability. Now, the expected wage payment per worker depends on $q$, as monitoring at a higher rate lowers the gain to shirking. The moderator faces a trade-off: increasing $q$ reduces efficiency but lowers the rents captured by workers. The optimal monitoring probability is strictly positive, which leads to inefficient duplication of tasks. At the optimum, the moderator's expected payment per image is approximately 0.10, which is significantly higher than in the full commitment contract. Once again, *bundling* is beneficial; under partial commitment, it significantly reduces both the rents captured by workers and the inefficiency of monitoring.[7]

[7]The calculations from this section are available from the authors upon request.

# 3 Model

## 3.1 Set-up

A firm is faced with a countably infinite stream of independent and identical tasks $t = 1, 2, ...$ and can delegate tasks to a countably infinite pool of workers $i = 1, 2, ....$ 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 a worker's decision problem on a single task. The following convention is maintained: objects pertaining to tasks are subscripted and objects pertaining to workers are superscripted.[8]

**The Task.** Each task $t$ has an unknown state $\omega_t$ drawn from finite set $\Omega$ with common prior belief $\pi \in \Delta(\Omega)$. The firm can hire workers to learn about the state. Let $n_t \in \{1, 2\}$ denote the number of workers hired for task $t$ when the firm delegates the task, and let $\mathcal{I}_t$ be the ordered set containing the identities of these workers.[9] Workers do not observe the number or identity of other workers hired for a task.

A worker $i$ assigned to task $t$ chooses an *unobservable* effort level $e_t^i \in \{0, 1\}$. Exerting effort ($e = 1$) perfectly reveals the state and is costly; we normalize this cost to $c = 1$. No effort ($e = 0$) yields no information about the state and is costless. Let $s_t^i \in \mathcal{S} := \Omega \cup \{\emptyset\}$ be the information worker $i$ observes about task $t$. After making an effort choice and observing information about the state, the worker sends a message to the firm, $m_t^i \in \mathcal{S}$. Information is *not verifiable*: the worker can send any message in $\mathcal{S}$, regardless of the information he observes.

Upon receiving messages from all hired workers on a task, the firm compiles a task-message profile, $(m_t^i)_{i \in \mathcal{I}_t} \in \mathcal{M}$, where $\mathcal{M} := \cup_{n=1}^2 \mathcal{S}^n$ is the set of all possible task-message profiles and chooses an action $A_t \in \Omega$. It receives a payoff of $v > 1$ if its action matches the state and zero otherwise.

[8]Where there is no risk for confusion, superscrips also denote powers.

[9]Restriction attention to hiring one or two workers is without loss of generality in the setting we study, as the firm would never find it optimal to hire more than two workers.

**The Contract.** The firm designs contracts to offer to workers. A contract $C$ comprises a set of tasks $\mathcal{J} = \{t_j\}_{j=1}^{J}$ for $J \geq 1$, a monitoring technology $Q$ and a wage structure $W$.

We refer to $J \equiv |\mathcal{J}|$ as the *job size*. A contract has *bundling* if it has multiple tasks, $J \geq 2$ and *maximal bundling* if it has the largest feasible job size (which, for exogenous legal or technological constraints faced by the firm, we assume to be finite).

Workers are monitored by comparing their messages to the messages of other workers assigned to the same task, which we refer to as *peer-monitoring*. The monitoring technology is the set of probability distributions over the number of workers hired for each task in a worker's job, $Q = (Q_j)_{t_j \in \mathcal{J}}$ where $Q_j \in \Delta(\{1, 2\})$. Let $\mathbf{q} = (q_1, ..., q_J)$ denote the monitoring rate for the contract, where $q_j = Q_j(2)$.

The wage structure depends on the worker's messages, as well as the messages from other workers for tasks in the worker's job. The firm aggregates the task-message profiles for all tasks in $\mathcal{J}$ into a report $r = (m_{t_j})_{t_j \in \mathcal{J}} \in \mathcal{R}$ for the worker, where $\mathcal{R} := \mathcal{M}^J$. A worker does not observe the messages of other workers, and therefore, does not observe his report (aside from his own messages). The firm reveals a report $\tilde{r} \in \tilde{R}(r)$ to the worker, where $\tilde{R} : \mathcal{R} \rightarrow \mathcal{R}$. That is, for each $r$, $\tilde{R}(r) \subset \mathcal{R}$ is the set of reports that the firm can reveal to a worker. Payment depends on the revealed report, $W : \mathcal{R} \rightarrow \mathbb{R}$.

The correspondence $\tilde{R}$ determines the verifiability of the observed report. We consider two sets of restrictions on $\tilde{R}$, which determines the commitment environment for the firm. Under *full commitment*, the firm's information is fully verifiable and it can commit to truthfully reveal the observed report.

**Definition 1** (Full Commitment)**.** *A firm can* fully commit *if for every report* $r \in \mathcal{R}$, $\tilde{R}(r) = \{r\}$.

Under *partial commitment*, the firm's information is partially verifiable: the content of messages from other workers is verifiable but the firm can fail to reveal messages. In other words, the firm can hide messages but it cannot fabricate messages. Define a report $\tilde{r} \subset r$ if $\tilde{r}$ can be constructed by removing messages from $r$.[10]

[10]For example, if $\Omega = \{0, 1\}$ and $J = 1$, report $(1, \cdot)$ is a subset of $(1, 1)$ and $(1, 0)$ but not a

**Definition 2** (Partial Commitment). *A firm can* partially commit *if for every report* $r \in \mathcal{R}$, $\tilde{R}(r) = \{\tilde{r} | \tilde{r} \subseteq r\}$.

Partial commitment is realistic if the messages from other workers are verifiable but the burden of proof lies on the firm to produce these messages. For example, the firm can produce evidence to distinguish a report with a mismatch from a report with a match, but cannot produce evidence to distinguish a report with no monitoring from a report with monitoring.

The commitment environment determines which wage structures are credible for the firm to offer. If $\tilde{r} \in R(r)$, it is not credible for the firm to pay a lower wage on $\tilde{r}$ than $r$, as the firm would simply reveal report $\tilde{r}$ when it observes $r$. Therefore, the firm faces a credibility constraint. For all $r \in \mathcal{R}$,

$$W(r) \leq W(\tilde{r}) \qquad \forall \tilde{r} \in \tilde{R}(r). \tag{CC}$$

We also restrict attention to wage structures in which workers are protected by limited liability.

**Condition 1** (Limited Liability). $W(r) \geq 0$ *for all* $r \in \mathcal{R}$.

Workers are identical and anonymous, so there is no strategic element to matching workers and job contracts, and tasks are identical, so there is no strategic element to assigning tasks to contracts. Given job size $J$, the firm fills the job with $J$ tasks and assigns it to the next available worker. Denote worker $i$'s contract by $C^i = (\mathcal{J}^i, Q^i, W^i)$, his effort profile for the job as $e^i = (e^i_j)_{t_j \in \mathcal{J}^i} \in \{0,1\}^{J^i}$, his signal profile as $s^i = (s^i_j)_{t_j \in \mathcal{J}^i} \in \mathcal{S}^{J^i}$ and his message profile as $m^i = (m^i_j)_{t_j \in \mathcal{J}^i} \in \mathcal{S}^{J^i}$, where $J^i = |\mathcal{J}^i|$ is the job size and $\mathcal{S}^{J^i} = \{\Omega \cup \emptyset\}^{J^i}$ is the set of all possible signal profiles.

Each set of tasks $\mathcal{J}$ has an underlying state vector $\boldsymbol{\omega} = (\omega_1, ..., \omega_J)$ of length $J = |\mathcal{J}|$. Let $\Pi_J(\boldsymbol{\omega}) \equiv \pi(\omega_1)\pi(\omega_2)...\pi(\omega_J)$ denote the probability of $\boldsymbol{\omega}$.

## 3.2 The Worker's Problem.

**Strategies.** Fix contract $C^i$ for worker $i$. The worker's strategy is a distribution over effort profiles $\sigma^i \in \Delta(\{0,1\}^{J^i})$ and a map from the set of signal profiles to

subset of $(0,0)$ or $(0,1)$, where $\cdot$ corresponds to no message from a second worker.

a distribution over message profiles $\mu^i : \mathcal{S}^{J^i} \to \Delta(\mathcal{S}^{J^i})$.[11] Abusing notation, let $\sigma^i_t$ be the probability that worker $i$ exerts effort on task $t$. Denote the set of all strategy profiles for player $i$ by $\Sigma^i$.

Two strategies play a prominent role in our analysis. Let $(\overline{\sigma}^i, \overline{\mu}^i)$ denote the strategy that exerts effort on all tasks and reports information truthfully, $\mu^i(s) = s \; \forall s \in \mathcal{S}^{J^i}$. When the worker exerts high effort on all tasks, the probability of signal $s \in \Omega^J$ is $\Pi_J(s)$. Let $(\sigma^i_0, m)$ denote the strategy that exerts effort on no tasks and reports message $m$.

**Payoffs.** The worker is risk-neutral and his payoff depends on the wage payment and cost of effort (recall this is normalized to one).[12] Given report $r^i$, revealed report $\tilde{r}^i \in \tilde{R}(r^i)$ and effort profile $e^i$, the payoff from the contract is

$$W^i(\tilde{r}^i) - \sum_{t \in \mathcal{J}^i} \mathbb{1}_{\{e^i_t = 1\}}.$$

Note it is independent of the state. Given contract $C^i$, a worker captures *rents* if his expected payoff is strictly positive.

**Incentives.** We are interested in equilibria in which workers exert high effort and report truthfully on all tasks, and the firm truthfully reveals the observed report. In such an equilibrium, a worker's incentives are governed by the probability distribution over his reports and the payment for each report, but are independent of the other workers' contracts. Conditional on other workers exerting effort and reporting truthfully, the distribution over reports depends on the worker's strategy, the monitoring distribution and the distribution over the state space. Let $g : \Sigma^i \to \Delta\mathcal{R}$ be the probability measure induced over reports when other workers play $(\overline{\sigma}^{-i}, \overline{\mu}^{-i})$, where $g(r|\sigma^i, \mu^i)$ is the probability of report $r$ when worker $i$ chooses strategy $(\sigma^i, \mu^i)$. The incentive constraint for worker $i$

[11] It is without loss of generality to define the message strategy as independent of realized effort, since each signal corresponds to a unique effort level.

[12] Normalizing the cost of effort is without loss of generality, given the payoff from matching the correct state $v$ can vary.

to play $(\overline{\sigma}^i, \overline{\mu}^i)$ is

$$\sum_{\mathcal{R}} W^i(r) g(r|\overline{\sigma}^i, \overline{\mu}^i) - J^i \geq \sum_{\mathcal{R}} W^i(r) g(r|\sigma^i, \mu^i) - \sum_{t \in \mathcal{J}^i} \sigma_t^i \qquad \forall (\sigma^i, \mu^i) \in \Sigma^i. \tag{IC}$$

A worker accepts a contract if and only if

$$\sum_{\mathcal{R}} W^i(r) g(r|\overline{\sigma}^i, \overline{\mu}^i) - J^i \geq 0. \tag{IR}$$

Note that under limited liability, (IR) is implied by (IC), as the worker earns a weakly positive payoff under the deviation to no effort on all tasks.

## 3.3 The Firm's Problem

**Strategies.** At the task level, the firm chooses how many workers to hire and an action choice, and at the worker level, the firm chooses how to design each contract. A task-strategy is a distribution over the number of workers to hire for each task, $\eta_t \in \Delta(\{0, 1, 2\})$, and a map from the set of message profiles to the set of distributions over the action choice for each task, $\alpha_t : \mathcal{M} \to \Delta(\Omega)$. We restrict attention to stationary task strategies $\eta$ in which $\eta_t = \eta$ for all $t$. This is without loss of generality, as every non-stationary task-strategy has a payoff-equivalent stationary task-strategy, given that tasks are symmetric and hiring multiple workers provides no additional information about the state.

**Monitoring Consistency.** The monitoring technology for a contract depends on the number of workers hired for each task in the contract. Therefore, the firm's choice of monitoring technology is linked to the firm's task-strategy $\eta$. The monitoring consistency condition requires that the firm's monitoring technology is consistent with the number of worker-tasks generated by the firm.

**Condition 2** (Monitoring Consistency)**.** *Given stationary task-strategy $\eta$ with*

$\eta(1) = 1 - p$ *and* $\eta(2) = p$*, monitoring rate* $\mathbf{q}$ *is consistent if*

$$\frac{1}{J}\sum_{j=1}^{J} q_j \quad = \quad \frac{2p}{1+p}. \tag{1}$$

Given a monitoring rate $\mathbf{q}$, there is a unique stationary task-strategy $\eta$ that satisfies Condition 2 (the converse is not true). Therefore, we can restrict attention to characterizing the firm's optimal contract, and the stationary task-strategy that is consistent with this contract follows from monitoring consistency. The following example provides intuition for this condition.

**Example 1.** Suppose that the firm hires two workers for a task with probability $p = 1/2$. In expectation this generates three worker-tasks for every two tasks and two of these three worker-tasks are assigned to two workers. From a worker's perspective, the probability that the task has been assigned to a second worker is 2/3 and the probability that it has only been assigned to the current worker is 1/3. Consistency requires monitoring rate $q = 2/3$. More generally, if the firm offers a single contract with the same monitoring rate for all tasks, consistency requires monitoring rate $q = \frac{2p}{1+p}$.

**Payoffs.** The firm's payoff on a task depends on whether its action matches the realized state of the world and the payments to workers. The payoff from delegating task $t$ to set of workers $\mathcal{I}_t$, offering contracts $\{C^i\}_{i\in\mathcal{I}_t}$, receiving reports $\{r^i\}_{i\in\mathcal{I}_t}$ and choosing action $A_t$ is

$$v \times \mathbb{1}_{\{A_t=\omega\}} - \sum_{i\in\mathcal{I}_t} W^i(r^i)/J^i,$$

where $W^i(r^i)/J^i$ is the per-task payment for a worker hired for task $t$. This ensures that payments are not double-counted across tasks.[13]

[13]It may seem more natural to define the firm's payoff on a self-contained block of tasks and workers, in which all workers in the block are assigned to tasks in the block and vice versa. For any monitoring technology, job size and monitoring technology with $\eta(n) \in \mathbb{Q}$ for $n = 0, 1, 2$, where $\mathbb{Q}$ is the set of rational numbers, it is possible to form such a block. Choose $T$ tasks such that the number of worker-tasks, $T\sum_{n=1}^{2}\eta(n)n$, and the number of workers, $I = T\sum_{n=1}^{2}\eta(n)n/J$, are integers. For example, if $\eta(1) = \eta(2) = 1/2$ and $J = 10$, setting

To highlight when bundling yields a strict improvement for the firm, we assume that the firm places lexicographic weight on the job size. If two contracts result in the same payoff per-task for the firm, it chooses the contract with the smaller job size. This is for expositional convenience.

**Contract Design.** The firm offers a set of contracts $\mathcal{C}$ to maximize its expected per-task payoff, subject to the workers' incentive and individual rationality constraints and the firm's limited liability, monitoring consistency and credibility constraints. The firm solves

$$\max_{(\alpha,\mathcal{C})} \mathbb{E}\left[v \times \mathbb{1}_{\{A_t=\omega\}} - \sum_{i\in\mathcal{I}_t} W^i(r^i)/J^i\right] \tag{*}$$

subject to IC and IR for all $i$, Conditions 1 and 2 and CC. In this framework, the decision to not delegate tasks to workers corresponds to setting $J^i = 0$ and $W^i(r) = 0$ for all $r \in \mathcal{R}$ and $i$ and $\mathcal{I}_t = \emptyset$ for all $t$.

We define several types of monitoring in the context of the optimal contract. Let $\mathcal{Q}(\varepsilon) = \{Q | q_j \geq \varepsilon \ \forall j\}$ be the set of monitoring structures with a monitoring rate of at least $\varepsilon$ on each task. We say that *virtual monitoring* is optimal if, for any $\varepsilon > 0$, when the firm is restricted to the set of contracts with monitoring structures in $\mathcal{Q}(\varepsilon)$, the optimal contract sets $q_j = \varepsilon$ for all $j = 1, ..., J$. *Stochastic monitoring* is optimal if there exists an $\varepsilon > 0$ such that when the firm is restricted to the set of contracts with monitoring structures in $\mathcal{Q}(\varepsilon)$, the optimal contract sets $q_j > \varepsilon$ for all $j = 1, ..., J$.

## 3.4 Benchmarks

**First-best solution.** Suppose that effort is observable. The surplus from delegating a task to a worker equals the value of learning the correct state minus

$T = 20$ generates 30 worker-tasks to be completed by $I = 3$ workers. The firm's payoff on this block is

$$v \sum_{t=1}^{20} \mathbb{1}_{\{A_t=\omega\}} - \sum_{i=1}^{3} W^i(r^i)$$

Maximizing the expected payoff per-task is equivalent to maximizing the expected payoff per block. Thus it is valid to define the firm's objective function in terms of the per-task payoff.

the cost of effort, $v - 1$. If the firm does not hire a worker, it chooses the action corresponding to the most likely state, which yields an expected surplus of $\bar{\pi}v$, where $\bar{\pi} = \max_\omega \pi(\omega)$ is the probability of the most likely state. Thus, the first-best task-strategy is to delegate to a single worker if $v - 1 > \bar{\pi}v$ and to not delegate otherwise. We assume that delegation is efficient.

**Assumption 1** (Delegation is efficient)**.** *Assume* $v > 1/(1 - \bar{\pi})$.

Under this assumption, the first-best contract has no bundling ($J = 1$), no monitoring ($q_j = 0$ for all $j = 1, ..., J$) and pays the worker his cost of effort if he exerts effort. Workers do not earn rents.

**Unlimited Liability.** If there is no restriction on negative transfers, the firm can punish workers with arbitrarily severe punishments. In the optimal contract, the firm sets $W(r) < 0$ for reports $r$ that only occur under shirking profiles, ensuring that the expected payoff from any effort profile with shirking is negative.

**Lemma 1** (Unlimited Liability)**.** *Under both partial and full commitment, the optimal contract has virtual monitoring, no bundling and punishes mismatches. For any* $\varepsilon > 0$*,*

$$W(r) = \begin{cases} 1 & r \in \Omega \cup \{(m^i, m^{-i}) \in \Omega^2 | m^i = m^{-i}\} \\ 1 - 1/(\varepsilon(1 - \overline{\pi})) & r \in \{(m^i, m^{-i}) \in \Omega^2 | m^i \neq m^{-i}\} \end{cases}$$

*when restricted to* $\mathcal{Q}(\varepsilon)$*.*

*Proof.* Omitted. □

Therefore, when there is no restriction on negative wages, neither unobservable effort nor partial commitment create inefficiencies. The optimal wage structure satisfies the worker's individual rationality constraint with equality and workers do not earn any rents. Approximate efficiency and zero rents are achieved independent of the job size, so there is no benefit to bundling.

# 4 The Optimal Contract

## 4.1 Towards the Optimal Wage Structure

With limited liability, using negative transfers to punish shirking is not possible. The firm must dissuade shirking by providing workers with rents. As in Shapiro and Stiglitz (1984), incentives are generated by the threat of losing these rents if caught deviating. Workers can guarantee themselves a positive expected payment by shirking on all tasks and sending a message to the firm *as if* they exerted effort and acquired signals. Therefore, any contract that satisfies the incentive constraint for high effort and truthful messages also satisfies individual rationality.

Fix the job size $J$. Under strategy $(\overline{\sigma}^i, \overline{\mu}^i)$, the set of messages that a worker sends with positive probability is $\Omega^J$, with $k \equiv |\Omega|^J$ elements. Partition the set of reports into $\mathcal{R}_M$, in which the worker sends a message $m \in \Omega^J$ and matches on all monitored tasks (this set includes profiles with tasks that are not monitored), and $\mathcal{R}_N$, in which a mismatch occurs on a monitored task or a worker sends a message $m \notin \Omega^J$.[14] When workers play $(\overline{\sigma}, \overline{\mu})$, any $r \in \mathcal{R}_N$ occurs with probability zero, since the probability of a mismatch or a message outside $\Omega^J$ is zero. For any deviation $(\sigma^i, \mu^i) \neq (\overline{\sigma}^i, \overline{\mu}^i)$, mismatches occur with positive probability. Therefore, it is never optimal to offer a positive transfer for a report in $\mathcal{R}_N$.

**Lemma 2.** *Under limited liability, the optimal wage structure satisfies $W(r) = 0 \; \forall r \in \mathcal{R}_N$.*

*Proof.* Using the partition over reports, rewrite the incentive constraint as

$$\sum_{r \in \mathcal{R}_N} W(r) \left[ g(r|\overline{\sigma}^i, \overline{\mu}^i) - g(r|\sigma^i, \mu^i) \right] + \sum_{r \in \mathcal{R}_M} W(r) \left[ g(r|\overline{\sigma}^i, \overline{\mu}^i) - g(r|\sigma^i, \mu^i) \right] \geq \left( J - \sum_{t \in \mathcal{J}^i} \sigma_t^i \right)$$

For any $r \in \mathcal{R}_N$, $g(r|\overline{\sigma}^i, \overline{\mu}^i) = 0 \leq g(r|\sigma^i, \mu^i)$ for all $(\sigma^i, \mu^i)$. Setting $W(r) > 0$ for $r \in \mathcal{R}_N$ lowers the left hand side of the incentive constraint, which is never optimal. □

Partition $\mathcal{R}_M$ into $\{R(m)\}_{m \in \Omega^J}$, where $R(m)$ is the set of possible reports

[14] For example, if $J = 1$ and $\Omega = \{0, 1\}$, then $\mathcal{R}_M = \{0, 1, 00, 11\}$ and $\mathcal{R}_N = \{\emptyset, 01, 10, \emptyset 0, \emptyset 1, 0\emptyset, 1\emptyset\}$.

that occur with positive probability when a worker sends message $m \in \Omega^J$ and workers play $(\overline{\sigma}, \overline{\mu})$. This is the set of reports generated by message $m$, excluding mismatch reports.[15] We say a wage structure is *simple* if it pays a wage of 0 on any report in $\mathcal{R}_N$ and for each message $m \in \Omega^J$, there exists a $w > 0$ such that for any $r \in R(m)$, the worker is either paid $w$ or 0.

**Definition 3** (Simple Wage Structure)**.** *A wage structure is* simple *if it can be represented as* $(\mathbf{w}, \boldsymbol{\rho})$*, where* $\mathbf{w} = (w_1, ..., w_k) \in \mathbb{R}^k_+$ *is a vector of wage payments and* $\boldsymbol{\rho} = (\rho_1, ..., \rho_k)$ *is a vector of sets of reports with* $\rho_j \subset R(m_j)$ *for each* $m_j \in \Omega^J$*, such that, given* $k = |\Omega|^J$*,*

$$W(r) = \begin{cases} w_j & \text{if } r \in \rho_j, \; j = 1, ..., k \\ 0 & \text{if } r \in \mathcal{R} \setminus \bigcup_{j=1}^{k} \rho_j. \end{cases}$$

In such a wage structure, the firm pays at most $k+1$ different amounts, even though there are $k(k+1)$ different reports in $\mathcal{R}$. Note that a simple wage structure satisfies limited liability and pays zero on $r \in \mathcal{R}_N$, as deemed optimal in Lemma 2.

## 4.2 The Optimal Contract: Full Commitment

Our first main result is a characterization of the optimal contract under limited liability and full commitment. We state the result in Theorem 1, and then outline the proof in a series of lemmas.

**Theorem 1.** *Under limited liability and full commitment, the optimal contract that delegates tasks to workers uses a virtual monitoring technology, maximal bundling and a simple wage profile in which the worker is paid a positive wage on reports in which he is monitored and matches on every task and zero on all other reports. For any* $\varepsilon > 0$*, set of feasible monitoring structures* $\mathcal{Q}(\varepsilon)$ *and maximum job size* $J$*, the optimal wage profile is*

$$W(r) = \begin{cases} \frac{J}{(|\Omega|^J - 1)\varepsilon^J} \frac{1}{\Pi_J(m_j)} & \text{if } r \in R_J(m_j), \text{ for each } m_j \in \Omega^J \\ 0 & \text{if } r \in \mathcal{R} \setminus \bigcup_{j=1}^{|\Omega|^J} R_J(m_j), \end{cases}$$

[15] For example, if $J = 1$, then $R(1) = \{1, 11\}$ and $R(0) = \{0, 00\}$.

*where $R_J(m)$ is the report in which an agent sends message $m$ and is monitored and matches on all $J$ tasks, $\Pi_J(m)$ is the probability of state profile $m$ and $|\Omega|^J$ is the cardinality of the state space for $J$ tasks. There exists a $\overline{v}$ such that for $v > \overline{v}$, it is optimal to delegate tasks to workers.*

The optimal wage structure takes the form of paying the worker a large positive wage when the worker is monitored and matches on all tasks. For example, if the worker is assigned 10 tasks, and is monitored and matches on all 10, the worker is paid a positive wage. But if the worker is monitored and matches on 9 or fewer tasks, or has any mismatches, the worker is paid nothing. As the monitoring rate $\varepsilon$ decreases, the worker is paid a higher wage with lower probability. For each message $m \in \Omega^J$, there is a single report on which the worker is paid a positive wage. This wage is inversely proportional to the probability $\Pi_J(m)$ of state profile $m$. Therefore, a less likely state profile corresponds to a higher wage payment. This can be viewed as a bonus for identifying less likely states. For example, if a worker is asked to review 10 images and 5% of images need to be flagged while the remaining 95% do not, the worker's wage is increasing in the number of images he flags.

The optimal contract uses virtual monitoring, and is therefore approximately efficient. However, the worker captures rents. This is because of limited liability, which guarantees a positive payoff to shirking in any contract that pays a positive wage on at least one report. In expectation, each worker is paid $\frac{|\Omega|^J}{|\Omega|^J-1}$ per-taskve wage on reports in which he matcheper-task cost of effort $c = 1$.

Therefore the firm strictly benefits from bundling and selects the maximum feasible job size. Bundling provides the firm with another option when negative punishments 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. This helps to relax the limited liability constraint and reduces the rents captured by workers.

The proof of Theorem 1 is outlined in the following series of lemmas.

**Optimal Wage Structure.** We first derive the optimal wage structure to enforce high effort and truthful reporting for a fixed monitoring rate $\mathbf{q}$ and job size $J$. Any wage structure that enforces high effort and truthful reporting must satisfy the following conditions:

1. For each $m \in \Omega^J$, there exists a report $r \in R(m)$ such that $W(r) > 0$.
2. For each task $t$, there exists a report $r$ with $W(r) > 0$ such that a worker is monitored and matches on task $t$.

If (1) does not hold, then a worker would never report $m$, and if (2) does not hold, then a worker would never exert effort and report truthfully on task $t$.

The characterization proceeds as follows. We calculate the optimal simple wage structure $(\mathbf{w}^*, \boldsymbol{\rho}^*)$ to deter deviations to the set of strategies $\{(\sigma_0, m_j)\}_{j=1}^k$ in which a worker shirks on all tasks and sends message $m_j \in \Omega^J$. We then show that $(\mathbf{w}^*, \boldsymbol{\rho}^*)$ also deters deviations to arbitrary strategy $(\sigma^i, \mu^i) \in \Sigma^i$, which makes it optimal in the class of simple wage structures. Finally, we show that $(\mathbf{w}^*, \boldsymbol{\rho}^*)$ is optimal in the set of all wage structures that satisfy limited liability and full commitment.

Given message $m \in \Omega^J$, let $R_J(m) \subset R(m)$ be the report where an agent is monitored and matches on all $J$ tasks. For example, if $J = 1$ and $\Omega = \{0, 1\}$, then $R_1(1) = \{11\}$.

**Lemma 3** (Optimal Wage Structure)**.** *Fix the monitoring rate* $\mathbf{q}$ *and job size* $J$*. Under limited liability and full commitment, the optimal wage structure to enforce* $(\overline{\sigma}, \overline{\mu})$ *is simple and takes the form* $(\boldsymbol{\rho}^*, \mathbf{w}^*)$*, where*

$$\begin{aligned} \boldsymbol{\rho}^* &= (R_J(m_1), ..., R_J(m_k)) \\ \mathbf{w}^* &= \frac{J}{(|\Omega|^J - 1)\overline{q}} \cdot (1/\Pi_J(m_1), ..., 1/\Pi_J(m_k)) \,. \end{aligned}$$

*where for each* $m_j \in \Omega^J$*,* $\Pi_J(m_j)$ *is the probability of state profile* $m_j$ *and* $\overline{q} = \prod_{t=1}^J q_t$ *is the probability of being monitored on all tasks. If report* $r \in R_J(m_j)$ *for some* $m_j$*, then* $W(r) = J/(|\Omega|^J - 1)\overline{q}\Pi_J(m_j)$ *and otherwise,* $W(r) = 0$*.*

Lemma 3 establishes that the optimal wage structure is simple and takes the form of paying the worker a positive wage only when he is monitored and matches on all tasks. The intuition is as follows. The likelihood ratio of the probability of a report under working compared to shirking and sending message $m$, $g(r|\overline{\sigma}^i, \overline{\mu}^i)/g(r|\sigma_0, m)$, is highest for the report in which the worker is monitored on all tasks. Therefore, paying a positive wage only on $R_J(m)$ generates the strongest incentives. The magnitude of the wage is driven by the ratio of the probability of a report in $R_J(m)$ under high effort to the probability under no effort and sending message $m$, $1/\Pi_J(m)$. This ensures that deviations to no effort and each message $m \in \Omega^J$ are equally profitable. Recall all proofs not presented in the text can be found in the Appendix.

**Optimal Monitoring and Bundling.** Next, we derive the optimal monitoring technology and bundling level. In a high effort and truthful reporting equilibrium, the firm learns the true state for each task under any contract that satisfies the worker's incentive constraint. Any incentive compatible contract provides the firm with the same information, and the firm chooses the optimal contract to minimize the cost of acquiring this information. Therefore, the optimal monitoring rate is determined by minimizing the firm's expected per-task wage bill.

Fix monitoring rate $\mathbf{q}$ and job size $J$ and let $\overline{q} = \prod_{t=1}^{J} q_t$ be the probability of being monitored on every task. Under the wage structure characterized in Lemma 3, a worker receives report $r \in \rho_j^*$ and is paid $w_j^* = J/(|\Omega|^J - 1)\overline{q}\Pi_J(m_j)$ with probability $p_j(\rho_j^*) = \overline{q}\Pi_J(m_j)$ for each $j = 1, ..., |\Omega|^J$, and is paid 0 with probability $1 - \overline{q}$. Therefore, the expected wage *per worker* is

$$\mathbf{p}(\boldsymbol{\rho}^*) \cdot \mathbf{w}^* = J\left(\frac{|\Omega|^J}{|\Omega|^J - 1}\right) \tag{2}$$

where $\mathbf{p}(\boldsymbol{\rho}^*) \equiv (p_1(\rho_1^*), ..., p_k(\rho_k^*))$ is the probability of each set of reports in $\rho^*$. Importantly, (2) is independent of $\mathbf{q}$.

The firm seeks to minimize the expected per-task wage, which depends on the expected number of workers hired for each task and the expected wage per worker. Suppose the firm offers a single contract with monitoring rate $\mathbf{q}$. Given monitoring consistency, the firm must use task strategy $\eta(2) = \left(\frac{1}{J}\sum_{j=1}^{J} q_j\right) / \left(2 - \frac{1}{J}\sum_{j=1}^{J} q_j\right)$. The expected number of workers hired is $1 + \eta(2)$. Therefore, the expected wage *per-task* is

$$\left(\frac{2}{2-\frac{1}{J}\sum_{j=1}^{J}q_j}\right)\left(\frac{|\Omega|^J}{|\Omega|^J-1}\right) \tag{3}$$

The firm's optimal monitoring rate and job size minimizes (3). Since (3) is increasing in each $q_j$ and decreasing in $J$, the firm chooses the smallest monitoring rate possible on each task – virtual monitoring – and bundles the maximum number of tasks – maximal bundling.

**Lemma 4** (Optimal Monitoring and Bundling)**.** *The optimal contract under limited liability and full commitment has virtual monitoring and maximal bundling.*

*Proof.* Follows immediately from the above characterization and Lemma 3. □

**Cost of Monitoring.** The per-task cost of monitoring at rate $q$ is endogenous; it depends on the wage structure and the monitoring rate,

$$c_F(q) = \left(\frac{q}{2-q}\right)\left(\frac{|\Omega|^J}{|\Omega|^J-1}\right).$$

Under virtual monitoring, the cost of monitoring is arbitrarily small. Therefore, the inefficiency that arises from limited liability is arbitrarily small when the firm has full commitment.

**Corollary 1** (Costly Monitoring)**.** *Under limited liability and full commitment, the cost of monitoring in the optimal contract is arbitrarily small: for any $\delta > 0$, there exists an $\varepsilon > 0$ such that $c_F(\varepsilon) < \delta$.*

*Proof.* Follows immediately from the optimality of virtual monitoring. □

## 4.3 The Optimal Contract: Partial Commitment

Our second main result is a characterization of the optimal contract under limited liability and partial commitment. We state the result in Theorem 2, and then outline the proof in a series of lemmas.

**Theorem 2.** *Under limited liability and partial commitment, the optimal contract that delegates tasks to workers has a symmetric stochastic monitoring technology,* *maximal bundling and a simple wage profile in which the worker is paid a positive wage on reports in which he matches on all monitored tasks and zero on all other reports. Given optimal monitoring rate* $\mathbf{q}^* = (q, ..., q)$ *and job size* $J$*, the optimal wage profile is characterized by*

$$W(r) = \begin{cases} \frac{J}{x_j(\sum_{j=1}^{k} \Pi_J(m_j)/x_j - 1)} & \text{if } r \in R(m_j), \text{ for each } m_j \in \Omega^J \\ 0 & \text{if } r \in \mathcal{R}_N, \end{cases}$$

*where* $R(m_j)$ *is the set of reports in which an agent sends message* $m_j$ *and matches on all monitored tasks,* $\Pi_J(m_j)$ *is the probability of state profile* $m_j$ *and*

$$x_j = \prod_{t=1}^{J} (1 - q + q\pi(m_{jt}))$$

*is the probability of a report in* $R(m_j)$ *when the worker shirks and sends message* $m_j = (m_{j1}, ..., m_{jJ})$*. There exists a* $\overline{v}$ *such that for* $v > \overline{v}$*, it is optimal to delegate tasks to workers.*

The optimal wage structure takes the form of paying the worker a positive wage when the worker matches on all tasks on which he is monitored. If the worker has any mismatches, he is paid nothing. The worker is paid the same wage for a message independent of how many tasks are monitored. For example, if the worker is assigned 10 tasks and sends message $m$, he is paid the same wage when he is monitored and matches on anywhere from 0 to 10 of these tasks. As the monitoring rate $q$ decreases, the worker is paid a higher wage for each message, but in contrast to full commitment, the probability of being paid this wage is independent of $q$. The wage for each message is inversely proportional to the probability $\Pi_J(m)$ of state profile $m$. Similar to full commitment, the worker receives a bonus for identifying less likely states.

The optimal contract uses stochastic monitoring, and is therefore inefficient. The size of the inefficiency depends on the dispersion of the distribution over states and the maximum job size. Once again, limited liability allows the worker to capture rents. These rents are higher under partial commitment relative to full commitment, as partial commitment places further restrictions on how the firm can punish the worker.

Under partial commitment, bundling plays a larger role than simply scaling up the firm's available rewards and punishments. It also lowers the per-task cost of monitoring by reducing the required wage on each message. This leads to a lower optimal monitoring rate and reduces the efficiency loss from stochastic monitoring. As before, bundling reduces the rents captured by workers by allowing the firm to link punishment across multiple tasks.

Workers are treated symmetrically, as are tasks within a worker's job. Intuitively, monitoring technologies with asymmetric monitoring probabilities either across tasks or workers make inefficient use of the firm's monitoring ability, since the expected wage payment at the optimal wage structure is convex in $q$. Fixing a task strategy, and therefore the expected number of workers per-task, the monitoring consistency condition pins down the feasible monitoring technologies. Given an asymmetric monitoring technology $\mathbf{q}$, there exists a symmetric monitoring technology $\mathbf{q}'$ that hires the same expected number of workers but pays a lower average expected wage per worker.

The proof of Theorem 2 is outlined in the following series of lemmas.

**Optimal Wage Structure.** Partial commitment translates to a condition on the monotonicity of $W$, which is not satisfied in the optimal wage structure for full commitment. The wage structure characterized in Lemma 3 only pays workers a positive wage on reports $\{R_J(m)\}$ in which a worker is monitored and matches on every task. Every report in $R(m)$ is a subset of the report $R_J(m)$. Therefore, under partial commitment, it is not credible for the firm to pay a positive wage for $R_J(m)$ and zero for other reports in $R(m)$ (recall the credibility constraint requires $W(r) \geq W(\tilde{r})$ for all $r \subset \tilde{r}$). More generally, it is not credible to condition payment on the realization of monitoring by paying higher wages for reports with more monitoring.

We first derive the optimal wage structure for a fixed monitoring rate $\mathbf{q}$ and job size $J$. The method to establish the optimal wage structure follows the same basic steps as the full commitment case.

**Lemma 5.** *Fix the monitoring rate* $\mathbf{q}$ *and job size* $J$*. Under limited liability and partial commitment, the optimal wage structure to enforce* $(\overline{\sigma}, \overline{\mu})$ *is simple and* *takes the form* $(\boldsymbol{\rho}^*, \mathbf{w}(\boldsymbol{\rho}^*))$*, where*

$$\begin{aligned} \boldsymbol{\rho}^* &= (R(m_1), ..., R(m_k)) \\ \mathbf{w}(\boldsymbol{\rho}^*) &= \frac{J}{\sum_{j=1}^{k} \Pi_J(m_j)/x_j - 1} \cdot (1/x_1, ..., 1/x_k) \end{aligned}$$

*such that for each* $m_j \in \Omega^J$*,* $x_j = \prod_{t=1}^{J}(1 - q_t + q_t\pi(m_{jt}))$ *is the probability of a report in* $R(m_j)$ *when the agent shirks and sends message* $m_j$ *and* $\Pi_J(m_j)$ *is the probability of state profile* $m_j$*. If* $r \in R(m_j)$ *for some* $m_j$*, then* $W(r) = J/x_j(\sum_{j=1}^{k} \Pi_J(m_j)/x_j - 1)$ *and otherwise,* $W(r) = 0$*.*

Lemma 5 establishes that the optimal wage structure is simple and takes the form of paying the worker the same wage on every profile in which the worker matches on any task on which he is monitored, regardless of the number of monitored tasks, and otherwise paying the worker a wage of zero. The wage for reports generated by message $m$ is inversely proportional to the likelihood that the worker matches on all monitored tasks when the worker shirks and reports $m$. For any monitoring rate $\mathbf{q}$ and job size $J$, this wage will be lower than the wage for the same message under full commitment, but the firm will pay this wage with higher probability.

The intuition for paying the same wage on all reports in $R(m)$ is as follows. Let $R_l(m) \subset R(m)$ be the set of reports in which the worker sends message $m$ and is monitored and matches on $l$ tasks. Since the difference between the probability of a report in $R_l(m)$ under working and shirking is increasing in $l$, the firm wants to pay a higher wage for reports in which the worker is monitored on more tasks. However, this is precisely what violates credibility. Therefore, the partial commitment requirement is binding and the firm cannot condition on the realization of monitoring to strengthen incentives.

**Optimal Monitoring and Bundling.** As in the case of full commitment, the optimal monitoring rate is determined by minimizing the expected per-task wage bill. Fix monitoring rate $\mathbf{q}$ and job size $J$. Under the wage structure characterized in Lemma 5, with probability $p_j(\rho_j^*) = \Pi_J(m_j)$ a worker sends message $m_j \in \Omega^J$ and is paid $w_j = J/x_j(\sum_{j=1}^{k} \Pi_J(m_j)/x_j - 1)$. The expected wage *per worker* is

$$\mathbf{p}(\boldsymbol{\rho}^*) \cdot \mathbf{w}^* = J \left( \frac{\sum_{j=1}^{k} \frac{\Pi_J(m_j)}{x_j}}{\sum_{j=1}^{k} \frac{\Pi_J(m_j)}{x_j} - 1} \right) \tag{4}$$

where $\mathbf{p}(\boldsymbol{\rho}^*) \equiv (p_1(\rho_1^*), ..., p_k(\rho_k^*))$ and $x_j$ is the probability of a report in $\rho_j^*$ when the worker deviates to always shirking and sending message $m_j$. Note that $x_j$ depends on $\mathbf{q}$ and $J$. Therefore, under partial commitment, the expected wage per worker depends on the monitoring rate. More frequent monitoring lowers each $x_j$, and therefore, the expected wage for a worker, and the expected wage approaches infinity as the monitoring rate becomes small.

Suppose the firm offers a contract with symmetric monitoring rate $\mathbf{q} = (q, ..., q)$ (we will establish that symmetric monitoring is optimal). Given monitoring consistency, the expected number of workers hired for each task is $\frac{2}{2-q}$. Therefore, the expected wage *per-task* is

$$\left( \frac{2}{2-q} \right) \left( \frac{\sum_{j=1}^{k} \frac{\Pi_J(m_j)}{x_j(q,J)}}{\sum_{j=1}^{k} \frac{\Pi_J(m_j)}{x_j(q,J)} - 1} \right). \tag{5}$$

where we write $x_j(q, J)$ to explicitly show the dependence of the expected per-task wage on $(q, J)$. The optimal monitoring rate and job size minimize (5). Since (5) is decreasing in $J$, the firm bundles the maximum number of tasks i.e. maximal bundling. However, now there is a trade-off between the expected number of workers hired and the expected wage per worker – the former is increasing in $q$ while the latter is decreasing in $q$. Minimizing (5) with respect to $q$ results in a strictly positive optimal monitoring rate i.e. stochastic monitoring.

**Lemma 6** (Optimal Monitoring and Bundling)**.** *The optimal contract under limited liability and partial commitment has symmetric stochastic monitoring and maximal bundling.*

*Proof.* Lemma 5 and the above characterization of the expected wage per-task establish the optimality of maximal bundling and stochastic monitoring in the class of contracts with symmetric monitoring. Lemma 11 (in the Appendix) establishes the optimality of symmetric monitoring. □

**Cost of Monitoring.** The equilibrium cost of monitoring is endogenously determined by the optimal wage structure and monitoring rate

$$c_P(q^*) = \left(\frac{q^*}{2-q^*}\right)\left(\frac{\sum_{j=1}^{k}\frac{\Pi_J(m_j)}{x_j(q^*)}}{\sum_{j=1}^{k}\frac{\Pi_J(m_j)}{x_j(q^*)} - 1}\right).$$

Under stochastic monitoring, the cost of monitoring is strictly positive. Therefore, the inefficiency that arises from limited liability is strictly positive when the firm has partial commitment.

**Corollary 2** (Costly Monitoring)**.** *Under limited liability and partial commitment, the equilibrium cost of monitoring is strictly positive.*

*Proof.* Follows immediately from $q^* > 0$ under limited liability and partial commitment. □

Under unlimited liability or full commitment, the equilibrium cost of monitoring is arbitrarily close to zero. Therefore, limited liability and partial commitment only lead to inefficiencies when they are jointly required.

## 4.4 Comparative Statics

**Full Commitment.** Increasing the job size strengthens the effect of bundling: shirking becomes less attractive since the worker needs to produce acceptable output for more tasks. Therefore, the per-task wage for each report and the expected wage payment per worker is decreasing in the maximum job size. As the prior becomes more extreme, a worker is more likely to match when shirking and sending the message corresponding to the more likely state, and therefore must be paid a higher wage to deter this deviation. Therefore, the wage for the more likely states and the expected wage payment per worker both increase. The following Corollary outlines comparative statics on the optimal wage.

**Corollary 3** (Wage)**.** *Let $W^*$ be the optimal wage structure and $\theta = \prod_{j=1}^{|\Omega|}\pi(\omega_j)$ be a measure of the dispersion of the distribution over the state space. Under limited liability and full commitment,*

1. *If* $\Pi_J(m)$ *is increasing (decreasing) in dispersion* $\theta$*, then* $W^*(r)$ *is decreasing (increasing) in dispersion* $\theta$ *for all* $r \in R_J(m)$*.*

2. *The expected wage payment per worker is decreasing in dispersion* $\theta$ *and decreasing in the maximum job size.*

*Proof.* The claims follow immediately from the wage derived in Lemma 3. □

The optimal monitoring probability decreases with the maximum job size. With a larger job size, it is optimal to hire fewer workers for each task and pay each worker a lower per-task wage. Taken together, the rents that the firm pays to a worker vanish as the job size becomes large.

**Corollary 4** (No Asymptotic Rents)**.** *Under limited liability and full commitment,*

$$\lim_{J\to\infty} \mathbf{p}(\rho^*) \cdot \mathbf{w}(\rho^*)^T / J = 1,$$

*Proof.* The claim follows immediately from a worker's expected wage per-task, which is $\mathbf{p}(\boldsymbol{\rho})^* \cdot \mathbf{w}^*/J = |\Omega|^{\bar{J}}/(|\Omega|^{\bar{J}} - 1)$. □

**Partial Commitment.** A more extreme prior increases the likelihood that a shirker will match when monitored, and the firm monitors at a higher rate to reduce the rents captured by workers. Monitoring and bundling are substitutes: when more tasks are assigned to a worker, each task can be monitored less intensively. This contrasts with the optimal contract with full commitment, in which the optimal monitoring technology (virtual monitoring) is independent of the prior and job size.

**Corollary 5** (Monitoring Rate)**.** *Let* $\theta = \prod_{j=1}^{|\Omega|} \pi(\omega_j)$ *be a measure of the dispersion of the state space. Under limited liability and partial commitment, the optimal monitoring rate* $q^*$ *is increasing in* $\theta$ *and decreasing in the maximum job size.*

*Proof.* The claims follow immediately from Lemma 5 and Theorem 2. □

Virtual monitoring is optimal asymptotically as the job size grows large. Therefore, the efficiency loss relative to the contractible effort benchmark vanishes and asymptotically, the optimal contract is approximately efficient. Similar comparative statics to Corollary 3 hold for the wage under partial commitment.

# 5 A Crowdsourcing Application

**Market Overview.** 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 given time (Caulfield 2011).[16] 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).[17]

The typical AMT contract specifies a set of tasks and a wage. Workers are paid a flat wage if their output is deemed acceptable and otherwise, their work is rejected and they are paid nothing. In general, the wage payment does not depend on the message a worker reports or how many other workers are hired to complete the same tasks. Hiring multiple workers has been highlighted as the most common method of quality assurance on AMT (Mason and Suri 2012). Obtaining multiple responses is cost-efficient for most tasks and tends to be effective for quality control (Snow, O'Connor, Jurafsky, and Ng 2008).

Field experiments carried out on AMT suggest workers respond to economic incentives in a predictable fashion and participants appear to treat their pay as performance dependent (Horton and Chilton 2010; Horton, Rand, and Zeckhauser

[16] 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%.

[17] 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). 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.

in press; Mason and Watts 2009; Paolacci, Chandler, and Ipeirotis 2010). Yet 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.

We present a simple model in order to formalize the optimal design of AMT contracts. The contract we characterize offers potential improvements over those currently used in practice.

**The Optimal AMT Contract.** Suppose the firm can either accept a worker's output and pay a flat wage or reject the worker's output and pay nothing. The firm can choose how to divide the set of reports into acceptance and rejection sets, and must pay the same transfer for all reports in each set. We refer to this as an AMT wage structure.

**Definition 4** (AMT Wage Structure)**.** *An AMT wage structure can be represented as* $(w, \rho)$*, where* $w \in \mathbb{R}_+$ *is a wage payment and* $\rho \subset \mathcal{R}$ *is a set of reports, such that*

$$W(r) = \begin{cases} w & \text{if } r \in \rho \\ 0 & \text{if } r \notin \rho. \end{cases}$$

Theorem 3 characterizes the optimal contract with an AMT wage structure.

**Theorem 3.** *Under limited liability and partial commitment, the optimal contract with an AMT wage structure has symmetric stochastic monitoring, maximal bundling and wage scheme*

$$W(r) = \begin{cases} \frac{J}{1-[1-q^*+\bar{\pi}q^*]^J} & \text{if } r \in \mathcal{R}_M \\ 0 & \text{if } r \notin \mathcal{R}_M \end{cases}$$

*where* $\bar{\pi} = \max_\omega \pi(\omega)$ *is the probability of the most likely state,* $\mathcal{R}_M$ *is the set of reports in which the worker sends a message* $m \in \Omega^J$ *and matches on all monitored tasks, and the optimal monitoring rate is*

$$q^* = \arg\min_q \left(\frac{2}{2-q}\right)\left(\frac{J}{1-[1-q+\bar{\pi}q]^J}\right). \tag{6}$$

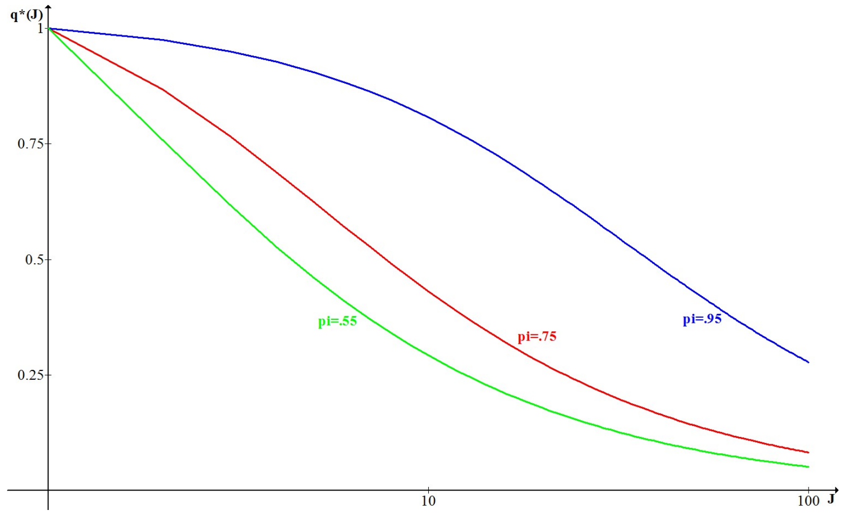


**Figure 1.** The optimal monitoring rate is decreasing in $J$.

This contract is inefficient, since the monitoring rate is strictly positive. Figure 1 plots the optimal monitoring rate for an example with a binary state space. As can be seen in the figure, the optimal monitoring rate is decreasing in the job size and increasing with the probability of the more likely state. As the job size becomes large, the optimal monitoring rate approaches virtual monitoring. Workers capture rents, since the wage is greater than the cost of effort, $1/(1-[1-q+\overline{\pi}q]^J) > 1$. The wage is also decreasing in the job size. Bundling reduces both the efficiency loss and the rents that the firm pays to a worker. As the job size grows large, both vanish. Figure Figure 2 plots this efficiency loss.

**Outline of Proof.** First we characterize the optimal AMT wage structure for a fixed monitoring rate $q$. It is obvious to see that the optimal AMT wage structure pays nothing on reports with mismatches, $W(r) = 0\ \forall r \in \mathcal{R}_N$ (this follows immediately from Lemma 2). We proceed by characterizing the optimal wage structure to prevent deviations to shirking on all tasks and reporting message $m_j \in \Omega^J$, and show this wage structure also prevents all other deviations. Given an AMT wage structure $(w, \rho)$, the incentive constraint for deviating to shirking

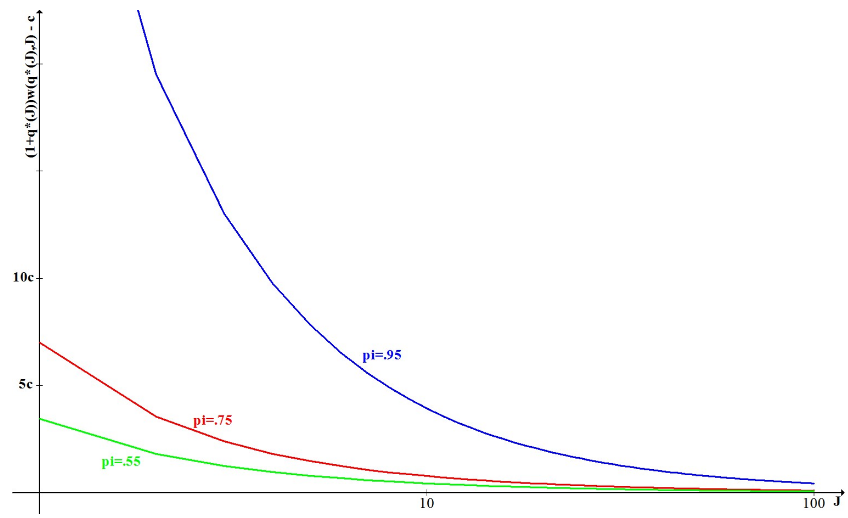


**Figure 2.** The efficiency loss is decreasing in $J$.

on all tasks and reporting message $m_j \in \Omega^J$ is

$$w \geq \frac{J}{p(\rho) - x_j(\rho)} \tag{7}$$

where $p(\rho) = g(\rho|\overline{\sigma}^i, \overline{\mu}^i)$ is the probability of a report in $\rho$ when the worker exerts high effort and reports truthfully and $x_j(\rho) = g(\rho|\sigma_0, m_j)$ is the probability of a report in $\rho$ when the worker deviates to shirking on all tasks and reporting $m_j$. Equation (7) cannot be satisfied with equality for each $m_j$ by paying different wages for different sets of reports. Therefore, for a given $\rho$, the deviation $m_j$ that maximizes $x_j(\rho)$ is binding.

The firm wants to choose $\rho$ to minimize the expected wage paid to a worker, $p(\rho)w$, subject to (7) holding for all $m_j$ and $\rho$ satisfying partial commitment. By analogous reasoning to Lemma 8, the optimal $\rho$ is the set of reports on which a worker matches on all monitored tasks, $\rho^* = \mathcal{R}_M$. When the worker exerts high effort and reports truthfully, the probability of a report in $\mathcal{R}_M$ is 1, and when the worker deviates to shirking on all tasks and reporting the message corresponding to the most likely state, the probability of a report in $\mathcal{R}_M$ is $(1 - q + \overline{\pi}q)^J$. Therefore, the optimal AMT wage structure to deter deviating to always shirking is to pay

$$w^* = \frac{J}{1 - [1 - q + \overline{\pi} q]^J} \tag{8}$$

on all $r \in \mathcal{R}_M$ and 0 elsewhere. As the monitoring rate increases, the profitability of shirking decreases and $w^*$ decreases.

Suppose the firm offers wage structure $(w^*, \rho^*)$ and consider a deviation to shirking and reporting the most likely state on $J - n$ tasks and exerting effort and reporting truthfully on $n$ tasks. The probability of receiving a report in $\mathcal{R}_M$ is $(1 - q + q\overline{\pi})^{J-n}$. This deviation isn't profitable if

$$w^* \geq \frac{J - n}{1 - (1 - q + q\overline{\pi})^{J-n}}.$$

The right hand side is decreasing in $n$. Since $(w^*, \rho^*)$ deters deviations to $n = 0$, it also deters all other deviations. Therefore, it is the optimal AMT wage structure.

As before, the firm chooses the monitoring rate that minimizes the expected payment per task. Given monitoring consistency, the expected number of workers hired for a task is $2/(2 - q)$ and the payment per worker is $w^*$. The optimal monitoring rate solves (6) and is strictly positive. It trades-off the efficiency loss from hiring more workers with the lower wage it pays these workers.

It is clear from (8) that bundling reduces the wage paid to workers by linking incentives across more tasks, and from (6) that bundling reduces the optimal monitoring rate. The firm bundles as many tasks as is feasible.

**The Gold Standard.** Another method of quality assurance, 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.[18] A worker's payment is based on whether his messages match on the known tasks. In our setting, the optimal peer-monitoring contract strictly outperforms the optimal gold standard contract.

**Theorem 4.** *The optimal Gold Standard contract is dominated by the optimal peer-monitoring contract characterized in Theorem 3.*

[18] For example, see www.crowdflower.com.

The intuition for the result is as follows. Suppose the firm seeds a worker's job with tasks for which it already knows the state. 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. A worker's job is judged on the basis of the worker's performance on the subset of known tasks. Payment is provided if and only if the worker performs satisfactorily on all $n$ known tasks. A worker must be compensated for the cost of effort on all $J$ tasks, as well as receive whatever rents are required to dissuade shirking. Let $w_G$ denote the optimal wage. In equilibrium, the firm pays each worker $w_G$ and learns the state of $J - n$ new tasks. If $2J$ tasks are assigned to two workers, the cost is $2w_G$ while the benefit is $2(J - n)$.

Alternatively, the firm could have the two workers overlap on $n$ of the $J$ assigned tasks (i.e. peer-monitoring). The wage that incentivizes effort is the same, so the expected wage bill is the same as in the Gold Standard contract, $2w_G$. But now the firm is learning the state for $J + J - n = 2J - n$ new tasks. The benefit is strictly higher than in the Gold Standard contract, as the same wage bill is spread over a greater number of new tasks. Gold standard contracts are inefficient because they monitor workers independently, compared to the optimal peer-monitoring contract, which jointly monitor workers.

However, the gold standard is useful for robust contract design. Any peer-monitoring contract always has a babbling equilibrium in which no workers exert effort. Inserting known tasks into a worker's job can eliminate this undesirable equilibrium. The firm can use a combination of peer-monitoring and the gold standard to design a the optimal robust contract.

# 6 Discussion and Conclusion

The insights from Section 4 also apply to multidimensional tasks. Suppose an agent is assigned a task with multiple components. For example, a worker must complete a tax return with additional schedules for each source of non-wage income. Theorems 1 and 2 suggests that the Internal Revenue Service should jointly monitor all tax forms and impose the harshest possible penalty upon uncovering any irregularities.

An entertaining example of bundling incentives on multidimensional tasks 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 [...] 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 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).

The model is also robust to several other extensions, including asymmetric firm payoffs and almost perfect signals.

**Asymmetric Firm Payoffs.** Suppose the firm's payoffs are asymmetric across states and actions. Let $\alpha_\omega \geq 0$ be the firm's payoff from matching the state when the state is $\omega$ and $\beta_\omega \leq 0$ be the firm's payoff from failing to match state $\omega$. For example, in an image screening task, $\alpha_1$ ($\beta_1$) corresponds to permitting (prohibiting) the sharing of harmless content and $\alpha_0$ ($\beta_0$) represents removing (failing to remove) an objectionable item. The structure of the optimal contract is unchanged – only the firm's participation constraint changes. With payoffs $\alpha_\omega$ and $\beta_\omega$, the firm's expected payoff from selecting $A_t = 0$ ($A_t = 1$) without hiring any workers is $(1-\pi)\alpha_0 + \pi\beta_1$ ($(1-\pi)\beta_0 + \pi\alpha_1$). The value of the proposed contract must exceed both of these values in order for the firm to participate.

**Noisy Signals.** In Section 3, effort perfectly reveals the state for each task. Suppose that signals are imperfect but sufficiently precise so that in the contractible effort benchmark the firm still wants to hire a single worker. In the perfect signal model, workers always produce matching messages and obtain payment with certainty in equilibrium. This is not the case with imperfect signals. Even if all workers exert effort, workers sometimes send different messages for the same task.

There are three levers the firm can use to introduce leniency into the contract. It can set a more tolerant match rate for reports, it can bundle fewer tasks in each job or it can 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 receive a positive wage. For sufficiently precise signals, the firm finds it optimal to monitor less frequently, and otherwise the optimal contract is similar to Section 4. Incentive concerns push the firm to inefficiently hire multiple workers for a single task. The firm now receives an additional (small) learning benefit from hiring these additional workers. The firm does not pay on mismatched reports and bundles as many tasks together as possible.

In 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. But 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, and 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.

# References

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

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

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

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

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

Frei, B. (2009): "Paid crowdsourcing: Current state & progress toward mainstream business use," *Smartsheet White Paper*.

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

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*.

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.

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.

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

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

Mirrlees, J. A. (1999): "The theory of moral hazard and unobservable behaviour: Part I," *The Review of Economic Studies*, 66(1), 3–21.

oDesk (2012): "oDesk at a Glance," *https://www.odesk.com/info/about/*.

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

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

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.

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.

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

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

# A Appendix

**Proof of Lemmas 3 and 5.** The proof follows from Lemmas 7 - 10.

**Lemma 7.** *Fix a vector of sets of reports* $\boldsymbol{\rho} = (\rho_1, ..., \rho_k)$*, monitoring rate* $\mathbf{q}$ *and job size* $J$*. The optimal simple wage structure to deter deviations from high effort and truthful reporting to strategies* $\{(\sigma_0, m)\}_{m \in \mathcal{S}^J}$ *is* $\mathbf{w}(\boldsymbol{\rho}) = (w_1, ..., w_k)$*, where*

$$w_j(\rho_j) = \left( \frac{J}{\left( \sum_{l=1}^{k} \frac{p_l}{x_l} - 1 \right)} \right) \frac{1}{x_j},$$

$p_j \equiv g(\rho_j|\overline{\sigma}^i, \overline{\mu}^i)$ *is the probability of a report in* $\rho_j$ *under strategy* $(\overline{\sigma}^i, \overline{\mu}^i)$ *and* $x_j \equiv g(\rho_j|\sigma_0, m_j)$ *is the probability of a report in* $\rho_j$ *under strategy* $(\sigma_0, m_j)$ *for each* $m_j \in \Omega^J$*.*

The wage vector is driven by the ratio of the probability of a report in $\rho_j$ under high effort and the probability of a report in $\rho_j$ under shirking and sending message $m_j$. Note that $w_j$ only depends on $\mathbf{q}$ through the dependence of $p_j$ and $x_j$ on $\mathbf{q}$.

*Proof.* Fix an arbitrary $\boldsymbol{\rho} = (\rho_1, ..., \rho_k)$. Let $p_j = g(\rho_j|\overline{\sigma}^i, \overline{\mu}^i)$ be the probability of a report in set $\rho_j$ under strategy $(\overline{\sigma}^i, \overline{\mu}^i)$ and let $x_j = g(\rho_j|\sigma_0, m_j)$ be the probability of a report in set $\rho_j$ under the strategy $(\sigma_0, m_j)$. Note that $g(\rho_j|\sigma_0, m_l) = 0$ for all $l \neq j$. Let $\mathbf{p} = (p_1, ..., p_k)$ and $\mathbf{x} = (x_1, ..., x_k)$ be the corresponding vectors. The incentive constraint to prevent deviating to strategy $(\sigma_0, m_j)$ is

$$\mathbf{p} \cdot \mathbf{w}^T - J \geq x_j w_j$$

for each $j = 1, ..., k$. Define

$$\Delta = \begin{bmatrix} p_1 - x_1 & p_2 & ... & p_k \\ p_1 & p_2 - x_2 & & p_k \\ ... & & & ... \\ p_1 & p_2 & ... & p_k - x_k \end{bmatrix}$$

and $\mathbf{J} = (J, ..., J)$. Note the diagonal entries are negative, $p_j - x_j \leq 0$. Rewrite the set of incentive constraints as

$$\Delta \cdot \mathbf{w}^T \geq \mathbf{J}$$

and satisfy it with equality by setting $\mathbf{w}^T = \Delta^{-1} \cdot \mathbf{J}$, where $\Delta^{-1}$ is

$$\frac{1}{d}\begin{bmatrix} \prod_{j\neq 1} x_j - \sum_{j\neq 1} p_j \prod_{l\neq 1,j} x_l & p_2 \prod_{j\neq 2,1} x_j & ... & p_k \prod_{j\neq k,1} x_j \\ p_1 \prod_{j\neq 1,2} x_j & \prod_{j\neq 2} x_j - \sum_{j\neq 2} p_j \prod_{l\neq 2,j} x_l & & p_k \prod_{j\neq k,2} x_j \\ ... & & & ... \\ p_1 \prod_{j\neq 1,k} x_j & p_2 \prod_{j\neq 2,k} x_j & ... & \prod_{j\neq k} x_j - \sum_{j\neq k} p_j \prod_{l\neq k,,j} x_l \end{bmatrix}$$

and

$$d = \sum_{j=1}^{k} p_j \prod_{l\neq j}^{k} x_l - \prod_{j=1}^{k} x_j = \prod_{j=1}^{k} x_j \left( \sum_{j=1}^{k} \frac{p_j}{x_j} - 1 \right).$$

Summing each row and multiplying by $J/d$ yields:

$$\mathbf{w} \;=\; \frac{J}{d} \prod_{j=1}^{k} x_j \cdot (1/x_1, ..., 1/x_k) = \frac{J}{\left( \sum_{j=1}^{k} \frac{p_j}{x_j} - 1 \right)} \cdot (1/x_1, ..., 1/x_k)$$

For a given $\boldsymbol{\rho}$, this characterizes the vector of wages that satisfies the set of incentive constraints for deviations to $\sigma_0$ and sending a message $m \in \Omega^J$. Obviously, for any $\sigma^i$, sending a message $m \notin \Omega^J$ is never an optimal deviation, since $W(r) = 0$ for any report generated by $m \notin \Omega^J$. Therefore, $\mathbf{w}$ satisfies the set of incentive constraints for all deviations to strategies $\{(\sigma_0, m)\}$.

□

**Lemma 8.** *Fix a monitoring rate* $\mathbf{q}$ *and job size* $J$. *Given the simple wage vector* $\mathbf{w}(\boldsymbol{\rho})$ *characterized in Lemma 7, the optimal set of reports* $\boldsymbol{\rho}^*$ *to deter deviations to strategies* $\{(\sigma_0, m_j)\}_{j=1}^{k}$ *is:*

1. *Full Commitment:* $\boldsymbol{\rho}^* = (R_J(m_1), ..., R_J(m_k))$
2. *Partial Commitment:* $\boldsymbol{\rho}^* = (R(m_1), ..., R(m_k))$.

*Proof.* The optimal $\boldsymbol{\rho}$ is the set of reports that minimize the expected wage bill for a worker. The wage vector $\mathbf{w}$ and vectors of probabilities $\mathbf{p}$ and $\mathbf{x}$ depend on

$\boldsymbol{\rho}$; write $\mathbf{w}(\boldsymbol{\rho})$, $\mathbf{p}(\boldsymbol{\rho})$ and $\mathbf{x}(\boldsymbol{\rho})$ to capture this dependence. Given $\boldsymbol{\rho}$, the expected wage bill for a single worker is

$$\mathbf{p}(\boldsymbol{\rho}) \cdot \mathbf{w}(\boldsymbol{\rho})^T = \sum_{j=1}^{k} p_j(\rho_j) w_j(\rho_j) = J \frac{\left(\sum_{j=1}^{k} \frac{p_j(\rho_j)}{x_j(\rho_j)}\right)}{\left(\sum_{j=1}^{k} \frac{p_j(\rho_j)}{x_j(\rho_j)}\right) - 1},$$

where $p_j \leq x_j$. Minimizing the expected wage bill is equivalent to maximizing

$$\max_{\boldsymbol{\rho}} \sum_{j=1}^{k} \frac{p_j(\rho_j)}{x_j(\rho_j)},$$

where $1 \leq \sum_{j=1}^{k} \frac{p_j}{x_j} \leq k$.[19]

**Full commitment.** Suppose $\rho_j = R_J(m_j)$, the report where an agent is monitored and matches on all tasks. If a worker plays strategy $(\overline{\sigma}^i, \overline{\mu}^i)$, then the probability of a report in $\rho_j$ is the probability of message $m_j$, $\Pi_J(m_j)$, times the probability of being monitored on all tasks, $\overline{q} = \prod_{t=1}^{J} q_t$, times the probability of matching on all tasks conditional on playing $m_j$, which is 1. Therefore, $p_j(\rho_j) = \overline{q}\Pi_J(m_j)$. If a worker deviates to strategy $(\sigma_0, m_j)$, then the probability of a report in $\rho_j$ is the probability of message $m_j$, which is 1, times the probability of being monitored on all tasks, $\overline{q}$, times the probability of matching on all tasks conditional on playing $m_j$, $\Pi_J(m_j)$. Therefore, $x_j = \overline{q}\Pi_J(m_j)$. Hence, $p_j = x_j$. For all $r \in R(m_j)$ such that $r \notin R_J(m_j)$, $g(r|\overline{\sigma}^i, \overline{\mu}^i) < g(r|\sigma_0, m_j)$. Therefore, $p_j = x_j$ uniquely holds for $\rho_j = R_J(m_j)$ and $\frac{p_j(\rho_j)}{x_j(\rho_j)}$ is maximized at $\rho_j = R_J(m_j)$. Therefore, the optimal structure is $\boldsymbol{\rho}^* = (R_J(m_1), ..., R_J(m_k))$.

**Partial commitment.** Partial commitment requires that for all $r \in \rho_j$, if $r' \subset r$, then $r' \in \rho_j$. There must be at least one report $r \in \rho_j$, otherwise it is not possible to incentivize a worker to play message $m_j$. Let $r \in R_0(m_j)$. Then $r \subseteq r'$ for all $r' \in R(m_j)$. Given $\rho_j \neq \emptyset$, partial commitment requires $R_0(m_j) \in \rho_j$.

[19] The lower bound is derived from only including reports with no monitoring in $\boldsymbol{\rho}$, which yields $\sum_{j=1}^{k} \frac{p_j}{x_j} = \sum_{j=1}^{k} \binom{k}{j} \pi^{k-j} (1-\pi)^j = 1$, and upper bound from achieving $p_j = x_j$.

Suppose $\rho_j = R_0(m_j)$, the minimal set in which it is possible to satisfy partial commitment. Suppose $r_1 \in R_1(m_j)$. Let $m_{j1}$ be the element of $m_j$ which is monitored in report $r_1$. Then

$$\frac{p_j}{x_j} = \frac{g(R_0(m_j)|\overline{\sigma},\overline{\mu})}{g(R_0(m_j)|\sigma_0,m_j)} = \frac{\Pi_J(m_j)}{1} < \frac{g(r_1|\overline{\sigma},\overline{\mu})}{g(r_1|\sigma_0,m_j)} = \frac{\Pi_J(m_j)}{\Pi_1(m_{j1})}$$

Therefore,

$$\frac{p_j}{x_j} < \frac{g(R_0 \cup r_1|\overline{\sigma},\overline{\mu})}{g(R_0 \cup r_1|\sigma_0,m_j)},$$

and adding $r_1$ to $\rho_j$ would increase $p_j/x_j$ without violating partial commitment, given $r_1$ has a unique subset $r \in R_0(m_j)$, which is already in $\rho_j$.[20] Therefore $\rho_j' = R_0(m_j) \cup \{r_1\}$ is strictly preferred to $\rho_j$.

Let $R_2(m_j)|_{r_1} = \{r_2 \in R_2(m_j) : r_1 \subset r_2\}$. This is the set of reports that it is possible to add to $\rho_j'$ without violating partial commitment. To simplify notation, suppose $\mathbf{q}$ is symmetric with task monitoring rate $q$. Comparing $\rho_j'$ and $\rho_j'' = R_0 \cup R_1 \cup R_2|_{r_1}$,

$$\frac{g(\rho_j'|\overline{\sigma},\overline{\mu})}{g(\rho_j'|\sigma_0,m_j)} = \frac{\left[(1-q)^J + q\,(1-q)^{J-1}\right]\Pi_J(m_j)}{(1-q)^J + q\,(1-q)^{J-1}\,\Pi_1(m_{j1})}$$

$$\frac{g(\rho_j''|\overline{\sigma},\overline{\mu})}{g(\rho_j''|\sigma_0,m_j)} = \frac{\left[(1-q)^J + Jq\,(1-q)^{J-1} + (J-1)\,q^2\,(1-q)^{J-2}\right]\Pi_J(m_j)}{\left((1-q)^J + q\,(1-q)^{J-1}\,\Pi_1(m_{j1})\right)\left(1 + \frac{q}{1-q}\sum_{t=2}^{J}\Pi_1(m_{jt})\right)}$$

Since $\sum_{t=2}^{J}\Pi_1(m_{jt}) \leq (J-1)$,

$$\frac{g(\rho_j'|\overline{\sigma},\overline{\mu})}{g(\rho_j'|\sigma_0,m_j)} < \frac{g(\rho_j''|\overline{\sigma},\overline{\mu})}{g(\rho_j''|\sigma_0,m_j)}$$

and adding $R_1 \setminus \{r_1\} \cup R_2|_{r_1}$ to $\rho_j'$ would increase $p_j/x_j$ and still satisfy partial commitment. Therefore $\rho_j''$ is strictly preferred to $\rho_j'$. Analogous calculations establish the same property for asymmetric $\mathbf{q}$.

Similar logic establishes $\rho_j^{l+1} = \cup_{n=0}^{l} R_n \cup R_{l+1}|_{r_1}$ is strictly preferred to $\rho_j^l =$

[20] This inequality follows from if $a_1/b_1 < a_2/b_2$, then $a_1/b_1 < (a_1+a_2)/(b_1+b_2)$.

$\cup_{n=0}^{l-1} R_n \cup R_l|_{r_1}$ for all $l < J$ and satisfies partial commitment, where $R_l(m_j)|_{r_1} = \{r \in R_l(m_j) : r_1 \subset r\}$. Note $R_J(m_j)|_{r_1} = R_J(m_j)$ is a singleton. Therefore, $\rho_j^J = \cup_{n=0}^{J-1} R_n \cup R_J|_{r_1} = R(m_j)$ is preferred to any other set of reports containing $r_1$. This holds for all $r_1 \in R_1(m_j)$. Therefore, the optimal set of reports for message $m_j$ is $\rho_j^* = R(m_j)$, and the optimal structure is $\boldsymbol{\rho}^* = (R(m_1), ..., R(m_k))$.[21]

□

**Lemma 9.** *Fix a monitoring rate* $\mathbf{q}$ *and job size* $J$. *The optimal simple wage* $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ *to deter deviations to strategies* $\{(\sigma_0, m_j)\}_{j=1}^k$ *also deters deviations to any strategy* $(\sigma^i, \mu^i) \in \Sigma^i$. *Therefore,* $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ *is the optimal simple wage structure.*

*Proof.* Let $(\sigma_n, m)$ denote the strategy where player $i$ deviates to exerting effort and reporting truthfully on $n$ tasks and shirking and reporting message $m \in \Omega^{J-n}$ on $J - n$ tasks. Without loss of generality, assume the worker deviates on the first $J - n$ tasks. Given a contract with the simple wage $(\boldsymbol{\rho}^*, \mathbf{w}^*)$, the incentive constraint to deter a deviation to $(\sigma_n, m)$ is

$$\sum_{j=1}^{k} w_j^* \left[p_j(\rho_j^*) - g(\rho_j^* | \sigma_n, m)\right] \geq J - n.$$

Let $m_j|_x$ denote the first $x$ messages in a message profile.

**Full Commitment.** Under full commitment, $p_j(\rho_j^*) = \overline{q}\Pi_J(m_j)$ for state $m_j \in \Omega^J$ and $\sum_{j=1}^k w_j^* p_j(\rho_j^*) = J\left(\frac{|\Omega|^J}{|\Omega|^J - 1}\right)$. When the worker deviates, the probability

[21] It is more straight forward to establish that

$$\frac{g(R_0|\overline{\sigma}, \overline{\mu})}{g(R_0|\sigma_0, m_j)} < \frac{g(R_0 \cup R_1|\overline{\sigma}, \overline{\mu})}{g(R_0 \cup R_1|\sigma_0, m_j)} < ... < \frac{g(\cup_{n=0}^{l} R_n|\overline{\sigma}, \overline{\mu})}{g(\cup_{n=0}^{l} R_n|\sigma_0, m_j)} < ... < \frac{g(\cup_{n=0}^{J} R_n|\overline{\sigma}, \overline{\mu})}{g(\cup_{n=0}^{J} R_n|\sigma_0, m_j)}.$$

Therefore, $\rho_j' = \cup_{n=0}^{l+1} R_n$ is strictly preferred to $\rho_j = \cup_{n=0}^{l} R_n$ for all $l < J$. Adding $r \in R_{l+1}$ to $\rho_j$ will not violate partial commitment, as for any $r' \subset r$, $r' \in \rho_j$. However, under skewed priors, it is possible that adding a single report from $R_{l+1}$ is better than adding the whole set,

$$\frac{g(\rho_j \cup \{r\} | \overline{\sigma}, \overline{\mu})}{g(\rho_j \cup \{r\} | \sigma_0, m_j)} > \frac{g(\rho_j' | \overline{\sigma}, \overline{\mu})}{g(\rho_j' | \sigma_0, m_j)}$$

for some $r \in R_{l+1}$, so this doesn't establish the optimality of $\rho_j = \cup_{n=0}^{J} R_n$ across all sets of reports that satisfy partial commitment.

of a report in $\rho_j^*$ is $g(\rho_j^*|\sigma_n, m) = \overline{q}\Pi_J(m_j)$ for state $m_j \in \Omega^J$ if the first $J-n$ states match $m$, $m_j|_{J-n} = m$, and $g(\rho_j^*|\sigma_n, m_j) = 0$ for state $m_j \in \Omega^J$ if the first $J-n$ states do not match $m$, $m_j|_{J-n} \neq m$. Therefore, the incentive constraint for a deviation to $(\sigma_n, m)$ simplifies to

$$\left(\frac{J|\Omega|^J}{|\Omega|^J - 1}\right) - \sum_{\left\{m_j \in \Omega^J \middle| m_j|_{J-n} = m\right\}} \frac{J}{|\Omega|^J - 1} \geq J - n.$$

When a worker reports truthfully on $n$ tasks and shirks and plays message $m \in \Omega^{J-n}$ on the remaining tasks, there are $|\Omega|^n$ states in which the worker's message matches the state, $\left|\left\{m_j \in \Omega^J \middle| m_j|_{J-n} = m\right\}\right| = |\Omega|^n$. The incentive constraint simplifies to

$$\frac{J}{|\Omega|^J - 1}(|\Omega|^J - |\Omega|^n) \qquad \geq J - n \tag{9}$$

The RHS of(9) is decreasing linearly in $n$, while the LHS of (9) is decreasing and concave in $n$. At $n = 0$ and $n = J$, (9) is satisfied with equality. Therefore, (9) holds for all $n = 1, ..., J-1$. Intuitively, the cost (in lost wages) of deviating to shirking on $x$ tasks is concave in $x$, while the gain (in saved effort) is linear. Therefore, if it is profitable to shirk on one task, it is profitable to shirk on them all.

**Partial Commitment.** Under partial commitment, $p_j(\rho_j^*) = \Pi_J(m_j)$ for state $m_j \in \Omega^J$ and $w_j = J/(\sum_{j=1}^k \Pi_J(m_j)/x_j - 1)x_j$. When the worker deviates, the probability of a report in $\rho_j^*$ is $g(\rho_j^*|\sigma_n, m) = \Pi_J(m_j)$ for state $m_j \in \Omega^J$ if the first $J-n$ states match $m$, $m_j|_{J-n} = m$, and $g(\rho_j^*|\sigma_n, m_j) = 0$ for state $m_j \in \Omega^J$ if the first $J-n$ states do not match $m$, $m_j|_{J-n} \neq m$. When a worker reports truthfully on $n$ tasks and shirks and plays message $m \in \Omega^{J-n}$ on the remaining tasks, there are $|\Omega|^n$ states in which the worker's message matches the state, $\left|\left\{m_j \in \Omega^J \middle| m_j|_{J-n} = m\right\}\right| = |\Omega|^n$. The incentive constraint for a deviation to

$(\sigma_n, m)$ simplifies to

$$\left(\frac{J}{\sum_{j=1}^{k}\frac{\Pi_J(m_j)}{x_j}-1}\right)\left(\sum_{j=1}^{k}\frac{\Pi_J(m_j)}{x_j}-\sum_{\left\{m_j\in\Omega^J\middle|m_j|_{J-n}=m\right\}}\frac{\Pi_J(m_j)}{x_j}\right)\geq J-n. \tag{10}$$

where

$$\frac{\Pi_J(m_j)}{x_j}=\prod_{t=1}^{J}\frac{\pi(m_{jt})}{(1-q_t+q_t\pi(m_{jt}))}.$$

The RHS of(10) is decreasing linearly in $n$, while the LHS of (10) is decreasing and concave in $n$. At $n = 0$ and $n = J$, (10) is satisfied with equality. Therefore, (10) holds for all $n = 1, ..., J - 1$.

This establishes that $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ deters deviations to any pure strategy $(\sigma^i, m^i)$. Therefore, $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ deters deviations to any strategy $(\sigma^i, \mu^i) \in \Sigma^i$. Given that any wage structure must deter deviations to $\{(\sigma_0,m_j)\}_{j=1}^{k}$, and the optimal way to deter deviations to $\{(\sigma_0,m_j)\}_{j=1}^{k}$ also deters deviations to any strategy $(\sigma^i, \mu^i) \in \Sigma^i$, $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ is the optimal simple wage structure. □

**Lemma 10.** *The simple wage* $(\boldsymbol{\rho}^*, \mathbf{w}^*)$ *is optimal in the class of all wage structures that satisfy limited liability.*

*Proof.* Consider a wage structure where the firm does not pay the same amount on all reports with a positive wage in the set $R(m_j)$.

**Full Commitment.** It follows immediately from the optimal $\boldsymbol{\rho}^*$ that paying a positive wage on any reports $r \in R(m_j) \setminus R_J(m_j)$ will weaken the incentive constraint for deviations to strategies $\{(\sigma_0,m_j)\}_{j=1}^{k}$ and is therefore not optimal.

**Partial Commitment.** The firm would like to pay less on reports in $R(m_j) \setminus R_J(m_j)$ but the partial commitment constraint binds. Therefore, it is never optimal to pay more on such reports. □

**Lemma 11** (Monitor Symmetrically Across Tasks)**.** *For any monitoring technology* $\mathbf{q} = (q_1, ..., q_J)$ *with* $q_j \neq q_k$ *for some* $j, k$*, there exists another monitoring* *technology* $\tilde{\mathbf{q}} = (\tilde{q}, ..., \tilde{q})$ *that enforces high effort and truthful reporting for the same expected wage per worker, but results in hiring a lower expected number of workers.*

*Proof.* Fix $\mathbf{q} = (q_1, ..., q_J)$ and consider the deviation to always shirking and sending message $m_j = (m_{j1}, ..., m_{jJ}) \in \Omega^J$. Then

$$x_j = \prod_{t=1}^{J}(1 - q_t + q_t\pi(m_{jt})).$$

For any monitoring rate $\mathbf{q} = (q_1, ..., q_J)$, the firm can set a uniform monitoring rate $\tilde{\mathbf{q}} = (\tilde{q}, ..., \tilde{q})$ such that $x_j$ is the same. In other words, there exists $\tilde{q}$ such that

$$\prod_{t=1}^{J}(1 - \tilde{q} + \tilde{q}\pi(m_{jt})) = x_j.$$

Since the monitoring rate only influences the expected wage per worker through $x_j$, any two monitoring rates with the same $x_j$ result in the same expected wage per worker.

Under monitoring consistency, the expected number of workers hired for monitoring rate $\mathbf{q}$ is $n(\mathbf{q}) = 2/\left(2 - \frac{1}{J}\sum_{j=1}^{J} q_j\right)$ and the expected number of workers hired for monitoring rate $\tilde{\mathbf{q}}$ is $n(\mathbf{q}) = 2/(2 - \tilde{q})$. By the arithmetic-geometric mean inequality, the expected number of workers hired is lower under the symmetric monitoring rate. The firm cares about the expected wage bill. Therefore, within the set of contracts that lead to a given expected wage per worker, the firm picks the contract with the lowest expected number of workers hired – the symmetric monitoring contract. $\square$