OPPORTUNISM IN PRINCIPAL-AGENT RELATIONSHIPS WITH SUBJECTIVE EVALUATION
ONLINE APPENDIX

W. BENTLEY MACLEOD AND TECK YONG TAN

1. Malfeasance (For Online Publication Only)

This section incorporates *malfeasance* into the baseline model. By malfeasance, we refer to actions that the Agent can take to obtain a reward but do not have any benefit to the Principal (Becker and Stigler, 1974). We extend this model to allow for multi-tasking in the sense of Holmström and Milgrom (1991), and introduce the required additional ingredients. Next we consider the effect of malfeasance in the absence of guile. We provide a precise condition – the malfeasance-free condition in (1.6) – which distinguishes malfeasance from guile. Next, we describe how malfeasance affects our results on the two types of contracts considered in the main text. We end the discussion with an example in which guile and malfeasance each has no effect on efficiency by itself, but when they interact, the efficiency effect is of a first-order magnitude.

All missing proofs are found in Section 1.5.

1.1. **Additional Details to the Model.** Instead of exerting effort on just one task, the Agent now privately allocates effort to a number of tasks from the set $\{H, 1, 2, 3, ..., \bar{m}\} = \{H\} \cup \mathcal{M}$. Task H denotes the main task described in the baseline model. The effort vector is given by $\vec{\lambda} = [\lambda_H, \lambda_1, ..., \lambda_{\bar{m}}] \in [0,1]^{\bar{m}+1}$ where λ_τ is the effort exerted on task $\tau \in \{H\} \cup \mathcal{M}$. There are $\bar{m} + 2$ possible outcomes denoted by $o \in \{L\} \cup \{H\} \cup \mathcal{M}$. The probability of getting outcome $o \in \{H\} \cup \mathcal{M}$ is λ_o, and the probability of getting outcome L is $1 - \sum_{o \in \{H\} \cup \mathcal{M}} \lambda_o$.

Outcome $o = H$ is a successful production outcome that generates revenue $B > 0$ to the Principal, and outcome $o = L$ is a failed outcome that generates no revenue to the Principal. The other possible outcomes $o \in \mathcal{M}$ are "successful" malfeasance outcomes that generate no revenue to the Principal; hence $\mathcal{M}$ is the *malfeasance set* and any task $m \in \mathcal{M}$

Date: February 26, 2017.
Key words and phrases. Subjective Evaluation, Guile, Opportunism, Sales Contracts, Authority Contracts, Malfeasance.

corresponds to an act of malfeasance – costly actions that have no productive benefit, but might be rewarded via the incentive contract.

In particular, when a non-productive task in the malfeasance set $\mathcal{M}$ generates its "successful" outcome which has no value to the Principal, the Principal might still perceive it as valuable via the signal he receives. For example, a car that is sent in for routine maintenance may have no problem, but the mechanic may nevertheless perform unnecessary "repairs". Upon observing the amount of work done on his car, the car-owner might mistakenly believe that the mechanic has helped him avert certain disaster! The difference between malfeasance and shirking is that malfeasance still incurs costly effort by the Agent but it is intended only to affect the signals that the Principal receives, and thereby potentially enhances her income under the compensation contract. In contrast, shirking is a reduction in effort, rather than a reallocation of effort to another task.

The Agent's cost of exerting effort $\vec{\lambda}$ is $V\left(\sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau\right)$ where $V(\cdot)$ satisfies Assumption 1 and the Inada condition that $\lim_{\lambda\to 1} V(\lambda)=\infty$. To simplify notation, we also let $V\left(\vec{\lambda}\right)=V\left(\sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau\right)$. The signal-generating process continues to consist of a set of states $ts\in S^2$, where t is the signal to the Principal and s is the signal for the Agent. The probability of getting state ts under outcome o is still Γ^o_{ts} with $o\in\{L\}\cup\{H\}\cup\mathcal{M}$ now. The ex-ante unconditional probability of state ts is then:

$$\begin{aligned}\Gamma_{ts}\left(\vec{\lambda}\right)\equiv Prob\left[ts|\vec{\lambda}\right] &= \sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau\Gamma^\tau_{ts}+\left(1-\sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau\right)\Gamma^L_{ts}\\ &= \Gamma^L_{ts}+\sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau\hat{\Gamma}^\tau_{ts},\end{aligned}$$

where the marginal effect of effort λ_τ on the probability of state ts is:

$$\hat{\Gamma}^\tau_{ts} = \Gamma^\tau_{ts}-\Gamma^L_{ts}, \quad \forall\tau\in\{H\}\cup\mathcal{M}.$$

Due to the Inada condition, the Agent always chooses $\vec{\lambda}$ such that $\sum_{\tau\in\{H\}\cup\mathcal{M}}\lambda_\tau<1$, which ensures that we have a valid ex-ante probability distribution over states. Notice that since none of the malfeasance outcome generates any revenue, the first-best surplus continues to be $Surplus^{FB}$ as defined in (2.1), with first-best effort $\lambda_H=\lambda^{FB}$ and $\lambda_m=0$ $\forall m\in\mathcal{M}$.

The contracting game remains the same. A contract is formally defined as a triplet:

$$\psi=\left\{\vec{\lambda},\vec{c},\vec{w}\right\}\in\Psi\equiv[0,1]^{\bar{m}+1}\times\Re^{n^2}\times\Re^{n^2};$$

the difference now is that the effort obligation is from a higher dimension.

When the players truthfully report their signals, the Agent's optimal choice of $\vec{\lambda}$ is determined by the first-order conditions and associated complementary slackness conditions that for all $\tau \in \{H\} \cup \mathcal{M}$:

$$\hat{\vec{\Gamma}}^{\tau}\vec{w} - V'(\vec{\lambda}) \leq 0. \tag{1.1}$$

$$\lambda_{\tau}\left(\hat{\vec{\Gamma}}^{\tau}\vec{w} - V'(\vec{\lambda})\right) = 0, \tag{1.2}$$

Let $MB(\vec{w}) = \max_{\tau \in \{H\} \cup \mathcal{M}} \hat{\vec{\Gamma}}^{\tau}\vec{w}$. It is readily observed that a necessary condition for $\lambda_{\tau} > 0$ is that $\hat{\vec{\Gamma}}^{\tau}\vec{w} = MB(\vec{w})$. This has two implications. First, if $\hat{\vec{\Gamma}}^{H}\vec{w} < MB(\vec{w})$, then $\lambda_H = 0$. Second, if it is optimal for the Agent to choose $\lambda_H > 0$, then it is (weakly) optimal to set $\lambda_m = 0$ for all $m \in \mathcal{M}$. Since all malfeasance task $m \in \mathcal{M}$ generates no value to the Principal, malfeasance will never be part of an optimal contract. Hence we only have to consider contracts with $\lambda_H > 0$ and $\lambda_m = 0$ $\forall m \in \mathcal{M}$.[1]

Given these observations, and for notational conciseness, henceforth, when we write effort "λ", we are referring to an effort choice $\vec{\lambda} = [\lambda, 0, \ldots, 0]$; that is, $\lambda_H = \lambda$ and $\lambda_m = 0$ $\forall m \in \mathcal{M}$. This means $\Gamma_{ts}(\lambda)$ also denote $\Gamma^{L}_{ts} + \lambda\hat{\Gamma}^{H}_{ts}$, the probability of state ts under effort λ on only the productive task H, and analogously for the probability vector $\vec{\Gamma}(\lambda)$. Moreover, since all relevant contracts should be malfeasance-free, we denote a contract by $\psi = \{\lambda, \vec{c}, \vec{w}\}$ with the understanding that the effort vector obligation involved is $[\lambda, 0, ..., 0]$. The contract ψ will then have to be designed to make these choices incentive-compatible for the Agent.

In particular, from (1.1) and (1.2), a wage vector $\vec{w} \in \Re^{n^2}$ will induce effort obligation λ from the Agent if it satisfies the *incentive constraint for effort (ICE)*:

$$\hat{\vec{\Gamma}}^{H}\vec{w} = V'(\lambda), \tag{1.3}$$

which is equivalent to (3.3) in the main text, and the *incentive constraint for no-malfeasance (ICM)*:

$$\hat{\vec{\Gamma}}^{m}\vec{w} \leq V'(\lambda), \quad \forall m \in \mathcal{M}. \tag{1.4}$$

The Agent's participation constraint is similar to (2.8) in the main text:

$$U^{A}(\psi) = \vec{\Gamma}\left(\vec{\lambda}\right)\vec{w} - V\left(\vec{\lambda}\right) \geq U^{0}. \tag{1.5}$$

[1]More generally, if it is optimal to set $\lambda_H = 0$, then there is no need for performance pay and thus, no malfeasance.

1.2. **The Malfeasance Problem Without Subjective Evaluation.** We first put aside subjective evaluation and consider the case in which information is symmetric but imperfect. In other words, in this section, we remove steps (5) and (6) in the sequence of contracting game and instead assume that the signals t and s are publicly observed, and the terms in the contracts can be based directly on the signals instead of on the players' reports. We emphasize that information is symmetric by terming contracts here as *enforceable* contracts, as opposed to *self-enforcing* contracts where the parties' information is private and contracts then have to provide incentives for parties to also truthfully reveal their information.

An enforceable contract implements effort λ if it satisfies constraints PC (1.5), ICE (1.3) and ICM (1.4). The following proposition provides the condition for the implementation of an effort λ under the full support condition (FSC):

Proposition 1. *Suppose that the full support condition FSC holds at effort level λ. There exists an enforceable contract $\psi \in \Psi$, that implements λ if and only if the malfeasance-free condition (MFC) is satisfied:*

$$MF = \mathbb{H}^{++}\left(\vec{\Gamma}^H\right) \cap \left\{\cap_{m\in\mathcal{M}} \mathbb{H}^{+}\left(\vec{\Gamma}^H - \vec{\Gamma}^m\right)\right\} \neq \emptyset. \tag{1.6}$$

Any enforceable contract that implements λ satisfies $\vec{w} = \vec{c}$, together with constraints PC (1.5), ICE (1.3) and ICM (1.4).

The Malfeasance-Free Condition (MFC) in (1.6) is defined using the set of hyper-planes supporting $\vec{x} \in \Re^{n^2}$: $\mathbb{H}^{+}(\vec{x}) = \left\{\vec{y} \in \Re^{n^2} | \vec{x}^T\vec{y} \geq 0\right\}$. $\mathbb{H}^{++}$ is defined by replacing the weak inequality with a strict inequality.[2]

Because there is no need to provide incentives for parties to reveal their information in this case, conflict is unnecessary (i.e. $\vec{\delta} = \vec{0}$) in an enforceable contract as stated in the condition that $\vec{w} = \vec{c}$. Notice that the malfeasance-free condition (1.6) is independent of the effort choice. If malfeasance is not possible ($\mathcal{M} = \emptyset$), then a sufficient condition for the existence of a compensation scheme that induces effort is that the half-space (or blunt cone[3]) $P^H = \mathbb{H}^{++}\left(\vec{\Gamma}^H\right)$ is non-empty. This in turn is equivalent to requiring that $\vec{\Gamma}^H \neq \vec{0}$. Hence, in the absence of malfeasance, an enforceable contract can be written as long as there is even the slightest bit of information about the Agent's performance.

[2]The set $\mathbb{H}^{+}$ is called a half-space when $\vec{x} \neq \vec{0}$. This is because $\mathbb{H}^{+}(\vec{x}) \cup \mathbb{H}^{++}(-\vec{x}) = \Re^{n^2}$ and $\mathbb{H}^{+}(\vec{x}) \cap \mathbb{H}^{++}(-\vec{x}) = \emptyset$. More generally, the intersection of a number of hyper-planes forms a convex polytope and represents the set of vectors that satisfies a set of inequalities.

[3]A cone $P \subset \Re^{n^2}$ is any set with the feature that for all $\vec{x} \in P$, and any $\alpha > 0$, $\alpha\vec{x} \in P$. A cone is blunt if $\vec{0} \notin P$ and pointed if $\vec{0} \in P$.

When malfeasance is possible ($\mathcal{M} \neq \emptyset$), one needs to ensure that the compensation scheme does not encourage malfeasance. This requires that the wage vector $\vec{w}$ be in the no-malfeasance pointed cone $P^{MF} = \left\{\cap_{m\in\mathcal{M}}\mathbb{H}^+\left(\vec{\Gamma}^H - \vec{\Gamma}^m\right)\right\}$ as well. Any wage vector in P^{MF} sets the marginal reward to malfeasance lower than that for the productive effort λ_H.

This has some intuitive implications. Suppose that for each malfeasance task $m \in \mathcal{M}$, there is a state that is generated only by a "successful" malfeasance outcome m but never by H.[4] If there is also some state other than these that provides information on outcome H, then we can set punishments only in these states to make the marginal reward to malfeasance sufficiently low and hence deter malfeasance. This implies:

Corollary 1. *Suppose malfeasance is detectable; namely, for every $m \in \mathcal{M}$, there exists a state $ts_m \in S^2$ such that $\Gamma^H_{ts_m} = 0$ and $\Gamma^m_{ts_m} > 0$. If there is also a state $ts \neq ts_m$ $\forall m \in \mathcal{M}$ such that $\hat{\Gamma}^H_{ts} \neq 0$, then for any $\lambda \in [0,1]$, there exists an enforceable contract that implements λ.*

With regard to the first-best effort level, from Proposition 1, it is immediate that the MFC condition (1.6) provides conditions under which it can be implemented:

Corollary 2. *Suppose that the FSC holds at the efficient effort level λ^{FB}. There exists an enforceable contract that implements λ^{FB} if and only if the malfeasance-free condition (1.6) holds.*

Henceforth, we shall assume that condition MFC (1.6) is always satisfied so that any impossibility of contract formation or any inefficiency that arises is not (solely) due to the presence of malfeasance.

1.3. **Subjective Evaluation.** We bring back subjective evaluation and assume that signals t and s are private information of the Principal and the Agent respectively again.

1.3.1. *Sales Contracts and Guile.* In the presence of malfeasance, the PTR and ATR remains unchanged from the main text. Hence the set of sales contracts remains to be Ψ^S in (3.2). A good faith sales contract is defined as in Definition 2 with the addition of ICM (1.4).

The analogous guile-free constraint with (3.7) here is:

$$\vec{\Gamma}(\lambda)\vec{w} - V(\lambda) \geq \vec{\Gamma}\left(\vec{\lambda}^g\right)\Pi\vec{w} - V\left(\vec{\lambda}^g\right), \quad \forall\vec{\lambda}^g \in [0,1]^{\bar{m}+1}, \forall\Pi \in Z. \tag{1.7}$$

[4]FSC can still be satisfied by allowing outcome L to generate all possible states with strictly positive probability.

In contrast to (3.7), the set of deviating effort is of a higher dimension which includes deviations to engaging in malfeasance. Hence the guile-free constraint is more stringent in the presence of malfeasance, thus illustrating how malfeasance exacerbate guile. An optimal sales contract is defined as in Definition 3 with the more stringent guile-free constraint (1.7); when Π is the truthful-reporting constraint, (1.7) also implies ICM (1.4).

We can provide an analogous result to Proposition 2 and simplify the guile-free constraint (1.7) slightly. Define the optimal effort function $\Lambda(\cdot)$ as in (A.1). Under a wage vector $\vec{w}$ and reporting strategy Π, it is optimal for the Agent to load all effort onto task τ^* where:[5]

$$\tau^*(\Pi\vec{w}) = \arg\max_{\tau\in\{H\}\cup\mathcal{M}} \vec{\hat{\Gamma}}^{\tau}\Pi\vec{w}. \tag{1.8}$$

We can now define the new guile-function. Let

$$G(y) \equiv \Lambda(y)\,y - V(\Lambda(y)). \tag{1.9}$$

The analogous condition to (3.9) is thus:

$$\vec{\Gamma}(\lambda)\vec{w} - V(\lambda) \geq \vec{\Gamma}^{L}\Pi\vec{w} + G\left(\vec{\hat{\Gamma}}^{\tau^*(\Pi\vec{w})}\Pi\vec{w}\right) \quad \forall\Pi \in Z. \tag{1.10}$$

Let $\mathbb{G}^M(\lambda)$ be the set of wage vector $\vec{w}$ that satisfies condition (1.10).

Proposition 2. *Suppose that the full support condition FSC holds at λ. In the presence of malfeasance, there exists an optimal sales contract that implements λ if and only if the set $\mathbb{G}^M(\lambda)$ is non-empty.*

Proof. Omitted (see proof of Proposition 3). □

Since $\mathbb{G}^M(\lambda) \subset \mathbb{G}(\lambda)$, the existence condition under malfeasance is more stringent. Notice that Proposition 4 continue to hold here. In addition, in the presence of malfeasance, we have an additional result.

Proposition 3. *An optimal good-faith contract $\psi^{GF} = \left\{\lambda, \vec{p}^{GF}, \vec{w}^{GF}\right\}$ is not guile-free if there is a reporting strategy, Π, giving the Agent a payoff equal to his outside option, U^0, at the agreed upon effort λ, and:*

$$\exists m \in \mathcal{M} \quad \textit{such that} \quad \vec{\hat{\Gamma}}^{m}\Pi\vec{w} > V'(\lambda). \tag{1.11}$$

This result shows that malfeasance can increase the potential for guile, even if the contract is guile-free relative to the effort incentives on the productive task. Moreover,

[5] If there are more than one tasks that achieves the maximum, then pick the task with the lowest index with H indexing 0.

whenever the Agent has an incentive to choose malfeasance, the result implies that total effort will be *higher* than effort upon the productive task (since $\vec{\tilde{\Gamma}}^m \Pi \vec{w} > V'(\lambda)$). This result resonates well with the observation that in healthcare markets, there is a tendency for physicians to oversupply redundant services due to the additional compensation they receive for such services (see Dafny (2005) and Chandra et al. (2012)).

1.3.2. *Authority Contracts.* The constraints in the authority contractual relationship are unchanged with the addition of malfeasance, except that the contract now also has to satisfy ICM (1.4). Hence the set of authority contracts remains to be Ψ^A in (5.2), and the definition of an optimal authority contract is as in Definition 5 with the addition of ICM (1.4). For an exact same logic, the authority contract is always guile-free. We illustrate how the addition of malfeasance affects the existence of an optimal authority contract next:

Proposition 4. *Suppose that the full support condition FSC holds at λ. Then an optimal authority contract implementing λ exists if and only if the Strong Malfeasance-Free Condition SMFC (1.12) holds.*

$$SMFC = \mathbb{H}^{++}\left(\vec{\tilde{\gamma}}^H\right) \cap \left\{\cap_{m\in\mathcal{M}}\mathbb{H}^{+}\left(\vec{\tilde{\gamma}}^H - \vec{\gamma}^m\right)\right\} \neq \emptyset. \tag{1.12}$$

If the malfeasance set is empty, the SMFC reduces to $\mathbb{H}^{++}\left(\vec{\tilde{\gamma}}^H\right) \neq \emptyset$ which is equivalent to the condition in Proposition 7. In the presence of malfeasance, the Principal needs to have enough information to also distinguish outcome H from the malfeasance outcome m. Hence the malfeasance problem is a non-trivial problem in the authority contract, whereas it is always trivially guile-free. This thus illustrates that guile and malfeasance are different issues, although their interplay can be important, as shown in the example next.

1.4. **Example: The Interplay of Guile and Malfeasance.** The following example illustrates the interplay of guile and malfeasance. Suppose there are three possible signals: $S = \{U, A, E\}$, where the signals U, A and E represent "unacceptable", "acceptable" and "excellent" performance level respectively. Representing the signal distribution via a 3×3 matrix in the (t, s)-space with the convention of:

$$\Gamma^o = \begin{bmatrix} \Gamma^o_{UU} & \Gamma^o_{UA} & \Gamma^o_{UE} \\ \Gamma^o_{AU} & \Gamma^o_{AA} & \Gamma^o_{AE} \\ \Gamma^o_{EU} & \Gamma^o_{EA} & \Gamma^o_{EE} \end{bmatrix},$$

let the information structure be:

$$\Gamma^L = \begin{bmatrix} \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \eta & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{bmatrix} \quad \Gamma^H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \eta & 0 \\ 0 & 0 & 7\varepsilon \end{bmatrix}, \quad \Longrightarrow \quad \hat{\Gamma}^H = \begin{bmatrix} -\varepsilon & -\varepsilon & -\varepsilon \\ -\varepsilon & 0 & -\varepsilon \\ -\varepsilon & -\varepsilon & 7\varepsilon \end{bmatrix},$$

with the assumption that the effort obligation satisfies that $0 < \varepsilon < \eta < 7\varepsilon$.[6] Notice that all states, except for (A, A) and (E, E), indicate that the outcome is more likely to be the low outcome L. Hence effort incentives should only be provided when s is A or E.

1.4.1. *Good-Faith Sales Contract.* Consider the good-faith sales contract first. It can be verified that it is optimal to load all effort incentives at $s = E$. Hence the cost-price vector will be in the form of $p^S_U = p^S_A = p$, and $p^S_E = p + b$ with $b > 0$. To deter the Agent from always reporting $s = E$ to get the higher price, some conflict at state (U, E) and/or (A, E) is necessary: $\delta^S_{UE} > 0$ or $\delta^S_{AE} > 0$ or both. Conflicts are not necessary at all other states: $\delta^S_{ts} = 0\ \forall ts \neq UE, AE$. Hence the Agent's wage terms take the following form:

$$w = \begin{bmatrix} w_{UU} & w_{UA} & w_{UE} \\ w_{AU} & w_{AA} & w_{AE} \\ w_{EU} & w_{EA} & w_{EE} \end{bmatrix} = \begin{bmatrix} p & p & p + b - \delta^S_{UE} \\ p & p & p + b - \delta^S_{AE} \\ p & p & p + b \end{bmatrix},$$

Consider the Agent's ATR at $s = U$ first. To deter the Agent from reporting E, the contract must satisfy:

$$\begin{aligned} & (1-\lambda)\varepsilon(b - \delta^S_{UE}) + (1-\lambda)\varepsilon(b - \delta^S_{AE}) + (1-\lambda)\varepsilon b \leq 0 \\ \Longleftrightarrow \quad & \varepsilon(b - \delta^S_{UE}) + \varepsilon(b - \delta^S_{AE}) + \varepsilon b \leq 0. \end{aligned} \tag{1.13}$$

Next, for the Agent's ATR at $s = A$, to deter her from reporting E, the contract must satisfy:

$$\begin{aligned} & (1-\lambda)\varepsilon(b - \delta^S_{UE}) + \eta(b - \delta^S_{AE}) + (1-\lambda)\varepsilon b \leq 0 \\ \Longleftrightarrow \quad & \varepsilon(b - \delta^S_{UE}) + \tfrac{\eta}{1-\lambda}(b - \delta^S_{AE}) + \varepsilon b \leq 0 \end{aligned} \tag{1.14}$$

The rest of the ATR constraints are trivially satisfied.

The total expected conflict cost of the contract, which should be minimized in the optimal contract, is $(1-\lambda)\varepsilon(\delta^S_{UE} + \delta^S_{AE})$. It is readily verified that setting:

$$\delta^S_{AE} = b \quad , \quad \delta^S_{UE} = 2b \tag{1.15}$$

[6] In particular, this implies $7/14 > \epsilon > 1/14$.

will have both constraints (1.13) and (1.14) binding, and minimizes the total expected conflict cost.[7]

The value of b is determined by the ICE:

$$\vec{\tilde{\Gamma}}^{H}\vec{w} = V'(\lambda)$$
$$\iff \quad -\varepsilon\left(b-\delta^{S}_{UE}\right)-\varepsilon\left(b-\delta^{S}_{AE}\right)+7\varepsilon b = V'(\lambda)$$
$$\iff \quad b = \tfrac{V'(\lambda)}{8\varepsilon}, \tag{1.16}$$

and the value of p is set such that the Agent's participation constraint binds.

1.4.2. *Guile.* It turns out that if λ is high, the contract specified in (1.15) and (1.16) is also guile-free and is hence the optimal sales contract. This reason behind it is that under this contract, the Agent's effort incentives are (almost) never altered even when she uses a non-truthful reporting strategy. The effort incentives under Π is $\vec{\tilde{\Gamma}}^{H}\Pi\vec{w}$. If $\vec{\tilde{\Gamma}}^{H}\Pi\vec{w} = \vec{\tilde{\Gamma}}^{H}\vec{w}$, then the Agent's optimal effort choice under Π will still be the effort obligation λ, and the ATR constraints (which only deter lying after the Agent has exerted effort λ) will then suffice to ensure that the Agent cannot profitably deviate via guile.

To see why the Agent's effort incentives are unchanged under the contract in (1.15) and (1.16), consider what happens when the Agent mis-reports. If the Agent lies to E when $s=U$, then the change in effort incentives is:

$$[-\varepsilon p-\varepsilon p-\varepsilon p]-\left[-\varepsilon\left(p+b-\delta^{S}_{UE}\right)-\varepsilon\left(p+b-\delta^{S}_{AE}\right)-\varepsilon\left(p+b\right)\right]. \tag{1.17}$$

The terms in the first square bracket are the original effort incentives at $s=U$ when the Agent reports truthfully; the terms in the second square bracket are the new effort incentives at $s=U$ when the Agent lies and reports $s=E$. Next, if the Agent lies to E when $s=A$, then the change in effort incentives is:

$$[-\varepsilon p+0p-\varepsilon p]-\left[-\varepsilon\left(p+b-\delta^{S}_{UE}\right)+0\left(p+b-\delta^{S}_{AE}\right)-\varepsilon\left(p+b\right)\right]. \tag{1.18}$$

[7] Note that this is not the only good-faith sales contract. To characterize the set of conflicts for the good-faith sales contracts, first note that $b \le \delta^{S}_{AE}$ must hold. To see why, suppose, for a contradiction, that $b-\delta^{S}_{AE} > 0$. Then it must be true that $\delta^{S}_{UE} > b > 0$ to satisfy both constraints (1.13) and (1.14). Since $\eta > \varepsilon \implies \frac{\eta}{1-\lambda} > \varepsilon$, (1.14) is more stringent than (1.13). As the expected inefficiency of the contract is $(1-\lambda)\varepsilon(\delta^{S}_{UE}+\delta^{S}_{AE})$, $\frac{\eta}{1-\lambda} > \varepsilon$ then also implies that it is always more efficient to increase δ^{S}_{AE} and decrease δ^{S}_{UE} to satisfy (1.14) which then contradicts that $\delta^{S}_{UE} > 0$. Hence, it must be the case that $b-\delta^{S}_{AE} \le 0$ which then implies that (1.13) is more stringent than (1.14). Hence the set of $\{\delta^{S}_{AE}, \delta^{S}_{UE}\}$ that minimizes the expected inefficiency while satisfying (1.13) and (1.14) is not unique. Any:

$$\{\delta^{S}_{AE}, \delta^{S}_{UE}\}\text{-pair such that } \delta^{S}_{AE} \ge b \text{ and (1.13) binds,}$$

will suffice.

Similarly, the terms in the first square bracket are the original effort incentives at $s = A$ when the Agent reports truthfully; the terms in the second square bracket are the new effort incentives at $s = A$ when the Agent lies and reports $s = E$. Notice that under the contract in (1.15) and (1.16), both (1.17) and (1.18) are zero. Moreover, if the Agent lies to report A at $s = U$, or lies to report U at $s = A$, the effort incentives are clearly unchanged since the wages that the Agent gets remains unchanged. Hence, if Π is a reporting strategy in which the Agent is truthful at $s = E$, then from the discussion above, $\vec{\hat{\Gamma}}^H \Pi \vec{w} = \vec{\hat{\Gamma}}^H \vec{w}$ and the Agent's cannot profitably deviate via such a reporting strategy Π. Finally, suppose that the Agent lie at $s = E$ now. This implies that the Agent always gets p and her effort incentives $\vec{\hat{\Gamma}}^H \Pi \vec{w}$ is now 0, which means that she optimally puts in zero effort. It can be verified that p is negative when λ is sufficiently high,[8] and hence this is also not profitable.

1.4.3. *Malfeasance.* Suppose now that there is also a seemingly "harmless" malfeasance task m with distribution of states given by:

$$\Gamma^m = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Conditional on a "successful" malfeasance outcome m, the state is (E, U) with certainty; that is, the Principal believes that the performance is excellent while the Agent knows that it is unacceptable. This malfeasance task is "harmless" because under the guile-free incentive scheme derived in (1.15) and (1.16), the Agent will never engage in pure malfeasance (exert effort on m and then report her signal truthfully) since she has zero chance of getting the bonus at outcome m, while effort on m decreases the probability of reaching outcome L.

However, the Agent can benefit from engaging in a combination of malfeasance and guile. This involves the Agent reporting E more often after exerting effort on task m, to fool the Principal that the state is (E, E) when the true state is actually (E, U). Intuitively, such interplay of malfeasance and guile is an example of the Agent gaming the system to produce an excellent performance signal for the Principal and then lying ex-post to cover it

[8] p is set such that $\vec{\Gamma}(\lambda)\,\vec{w} - V(\lambda) = 0$.

$$\begin{aligned}\vec{\Gamma}(\lambda)\,\vec{w} - V(\lambda) =& p + (1-\lambda)\varepsilon\left(b - \delta^S_{UE}\right) + (1-\lambda)\varepsilon\left(b - \delta^S_{UE}\right) + 7\lambda\varepsilon b \\ =& p + \lambda V'(\lambda) - \frac{V'(\lambda)}{8} - V(\lambda)\end{aligned}$$

Hence $p = V(\lambda) - \lambda V'(\lambda) + \frac{V'(\lambda)}{8}$. The derivative of the right-hand side with λ is $V''(\lambda)\left[\frac{1}{8} - \lambda\right]$. Hence for λ sufficiently large, p will be negative.

up. The ability of the Agent to create a signal $t = E$ for the Principal via an activity (task m) that the Principal does not value makes it difficult to load incentives at E now. This illustrates the point that if a signal is easily gamed, it cannot be used to provide incentives. In such instances, some incentives have to be provided via a less informative signal that is harder to game. In this case, incentive has to be provided via the less informative signal A as well, which decreases the efficiency of the sales contract.

To illustrate the above intuition, denote $b_A \geq 0$ as the bonus that the Agent gets when she reports A; that is, $p_A^S = p + b_A$. As mentioned, it can be shown that in the absence of malfeasance, b_A will be optimally set to 0. Consider the additional malfeasance task now, which has marginal probabilities from effort:

$$\hat{\Gamma}^m = \begin{bmatrix} -\varepsilon & -\varepsilon & -\varepsilon \\ -\varepsilon & -\eta & -\varepsilon \\ 1-\varepsilon & -\varepsilon & 0 \end{bmatrix}$$

Consider the Agent exerting effort $\lambda_m > 0$ (to be determined) on task m and zero effort on task H, and then playing a reporting strategy Π of reporting E when she sees $s = U$, reporting truthfully when she sees $s = A$, and reporting U when she sees $s = E$.

The Agent's gain in expected payoff from doing so as opposed to adhering to the contract obligation can be expressed as:

$$\left[\vec{\Gamma}^L(\Pi - I)\vec{w}\right] + \left[\Big(\lambda_m V'(\lambda_m) - V(\lambda_m)\Big) - \Big(\lambda V'(\lambda) - V(\lambda)\Big)\right] \tag{1.19}$$

where λ_m is characterized by $V'(\lambda_m) = \vec{\hat{\Gamma}}^m \Pi \vec{w}$, the Agent's effort incentives on task m from her reporting strategy Π. For there to be no incentives for the Agent to engage in malfeasance with guile this way, (1.19) must be non-positive. The term in the first square bracket of (1.19) is:

$$\vec{\Gamma}^L(\Pi - I)\vec{w} = \varepsilon b,$$

which is strictly positive. Hence the term in the second square bracket of (1.19) must be strictly negative which then implies that λ_m must be strictly less than λ. Notice that:

$$\begin{aligned} V'(\lambda_m) &= \vec{\hat{\Gamma}}^m \Pi \vec{w} \\ &= -\varepsilon\left(b - \delta_{UE}^S\right) - \varepsilon\left(b - \delta_{AE}^S\right) + (1-\varepsilon)b - \varepsilon\left(b_A - \delta_{UA}^S\right) - \eta b_A - \varepsilon b_A \\ &= b + (\varepsilon - \eta)\, b_A, \end{aligned}$$

and $V'(\lambda) = \vec{\hat{\Gamma}}^H \vec{w} = (\varepsilon b_A + 8\varepsilon b)$. Hence

$$\begin{aligned} V'(\lambda_m) - V'(\lambda) &= (b_E + (\varepsilon - \eta)\, b_A) - (\varepsilon b_A + 8\varepsilon b) \\ &= (1 - 8\varepsilon)\, b - \eta b_A \\ &= (\eta - \varepsilon)\, b - \eta b_A \end{aligned}$$

where the last equality follows from $1 = \eta + 7\varepsilon$. With $\eta - \varepsilon > 0$, for λ_m to be strictly less than λ, it must then be the case that $b_A > 0$; incentives must thus be given at signal A as well.

1.5. **Proofs for Section 1.**

Proof of Proposition 1.

Proof. Suppose that $\vec{w}$ implements λ. Since $V'(\lambda) > 0$, ICE (1.3) implies that $\vec{\hat{\Gamma}}^H \vec{w} > 0$ and hence $\vec{w} \in \mathbb{H}^{++}\left(\hat{\Gamma}^H\right)$. ICE (1.3) and ICM (1.4) imply that $(\vec{\hat{\Gamma}}^H - \vec{\hat{\Gamma}}^m)\vec{w} \geq 0$ for all $m \in M$ and therefore $\vec{w} \in \mathbb{H}^{+}\left(\vec{\Gamma}^H - \vec{\Gamma}^m\right)$ $\forall m \in \mathcal{M}$. Thus we get the MFC as a necessary condition.

Conversely, suppose that the MFC condition is satisfied. Choose $\vec{w}^0 \in MF$. This implies that $\vec{\hat{\Gamma}}^H \vec{w}^0 > 0$, and hence we can choose $b > 0$ such that $b\vec{\hat{\Gamma}}^H \vec{w}^0 = V'(\lambda)$ so that $b\vec{w}^o$ satisfies ICE (3.3). Since $b > 0$, we also have $b\vec{w}^0 \in \left\{\cap_{m \in \mathcal{M}} \mathbb{H}^{+}\left(\vec{\Gamma}^H - \vec{\Gamma}^m\right)\right\}$ since this set is the intersection of positive cones. This implies that $b\vec{\hat{\Gamma}}^m \vec{w}^0 = b\vec{\hat{\Gamma}}^H \vec{w}^0 - b\left(\vec{\Gamma}^H - \vec{\Gamma}^m\right)\vec{w}^0 \leq V'(\lambda)$. $b\vec{w}^0$ thus satisfies ICM (1.4). Let $\vec{1}$ be a vector whose entries are all 1. Observe that $\vec{\Gamma}^\tau \vec{1} = 1$ $\forall \tau$ since these are probability vectors. Thus, for all $a \in \Re$, $(\vec{\Gamma}^H - \vec{\Gamma}^m)\vec{1}a = 0$, while $a\vec{\hat{\Gamma}}^H \vec{1} = 0$. Let $\vec{w}\,(a) = a\vec{1} + b\vec{w}^0$; $\vec{w}\,(a)$ thus satisfies both ICE (1.3) and ICM (1.4), and hence implements λ. Thus we have $U^A\,(\psi) = \vec{\Gamma}\,(\lambda)\,\vec{w} - V\,(\lambda) = a + b\vec{\Gamma}\,(\lambda)\,\vec{w}^0 - V\,(\lambda)\,,$ and we can choose a to satisfy the PC (1.5). ☐

Proof of Proposition 3.

Proof. Suppose Π gives the Agent a payoff equal to the outside option, U^0, at the agreed upon effort λ. If $\vec{\hat{\Gamma}}^H \Pi \vec{w} \neq V'\,(\lambda)$ then from Proposition 4 the contract is not guile free, and we are done. Consider now the case in which $\vec{\hat{\Gamma}}^H \Pi \vec{w} = V'\,(\lambda)$. Since information can be misrepresented at no cost, this implies $\vec{\Gamma}^L \Pi \vec{w} = \vec{\Gamma}^L \vec{w}$. Thus we have:

$$\begin{aligned}\vec{\Gamma}^L \vec{w} + \lambda \vec{\hat{\Gamma}}^H \vec{w} - V(\lambda) &= \vec{\Gamma}^L \Pi \vec{w} + \lambda \vec{\hat{\Gamma}}^H \Pi \vec{w} - V(\lambda), \\ &< \vec{\Gamma}^L \Pi \vec{w} + \lambda \vec{\hat{\Gamma}}^m \Pi \vec{w} - V(\lambda) \\ &\leq \vec{\Gamma}^L \Pi \vec{w} + \lambda^{m*} \vec{\hat{\Gamma}}^m \Pi \vec{w} - V(\lambda^{m*}),\end{aligned}$$

where m is the malfeasance task that provides strong malfeasance-effort incentives (1.11), and $\lambda^{m*} = \Lambda\left(\vec{\hat{\Gamma}}^m \Pi \vec{w}\right)$ where $\Lambda(\cdot)$ is as defined in (A.1). In other words, putting all her effort into malfeasance rather than the productive task makes the Agent strictly better off, and hence the contract is not guile-free. □

Proof of Proposition 4.

Proof. Suppose an optimal authority contract exists in the presence of malfeasance. Since $w_{ts} = p_t^A$, constraints ICE (1.3) and ICM (1.4) can be rewritten respectively as $\vec{\gamma}^H \vec{p}^A = V'(\lambda)$ and $\vec{\gamma}^m \vec{p}^A \leq V'(\lambda)\ \forall m \in \mathcal{M}$. An argument similar to the proof of Proposition (1) implies the SMFC (1.12) as a necessary condition.

Conversely, suppose SMFC (1.12) is satisfied and let $\vec{p}^0 \in SMF$. There exists $b > 0$ such that $b\vec{\gamma}^H \vec{p}^0 = V'(\lambda)$; $b\vec{p}^0$ thus satisfies ICE (1.3). Notice also that $b\vec{p}^0 \in SMF$ and hence satisfies ICM (1.4). Next choose $a \in \Re$ such that $\vec{p}^1 = a\vec{1} + \vec{p}^0$ satisfies the PC (1.5). Adding a unit vector $\vec{1}$ does not affect the satisfaction of the ICE and ICM constraints. Next, let $\bar{p} = \max_t p_t^1$ and set $\delta_{ts}^A = \bar{p} - p_t^1 \geq 0$. With this, we now have $c_{ts} = \bar{p}$ for all $ts \in S^2$ and hence PTR is satisfied trivially. This demonstrates that the feasible set is non-empty. Using the same argument as for Proposition 7, an optimal solution exists. □

References

Becker, G. S. and G. J. Stigler (1974). Law enforcement, malfeasance, and compensation of enforcers. *Journal of Legal Studies 3*(1), 1–18.

Chandra, A., D. Cutler, and Z. Song (2012). Who ordered that? the economics of treatment choices in medical care. In T. G. M. Mark V. Pauly and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume 2 of *Handbook of Health Economics*, pp. 397–432. Elsevier.

Dafny, L. S. (2005). How do hospitals respond to price changes? *The American Economic Review 95*(5), 1525–1547.

Holmström, B. and P. Milgrom (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization 7*, 24–52.

Department of Economics, Columbia University. E-mail: bentley.macleod@columbia.edu.
E-mail: teck.tan@columbia.edu