# Secret Contracts for Efficient Partnerships*

David Rahman† and Ichiro Obara
University of Minnesota

First Submission: November 6, 2005.
This Draft: June 24, 2008.

**Abstract**

By allocating different information to team members, secret contracts can provide better incentives to perform with an intuitive organizational design. For instance, they may help to monitor monitors, and appoint secret principals. Generally, secret contracts highlight a rich duality between detection and enforcement with linear transfers. On the one hand, disobedient deviations must be detectable to enforce a given outcome, but different behavior may be used to detect different deviations. On the other hand, disobedient deviations must be attributable, i.e., some individual can be identified as innocent, to provide incentives with budget balance.

*JEL Classification*: D21, D23, D82.
*Keywords*: secret contracts, partnerships, duality, private monitoring.

*This paper subsumes Rahman (2005a), which in turn is based on Chapter 3 of Rahman (2005b). Many thanks are owed to Harold Demsetz, Larry Jones, Narayana Kocherlakota, David Levine, Roger Myerson, Itai Sher, Joe Ostroy, Phil Reny, Joel Sobel, Bill Zame, co-editor Larry Samuelson and three anonymous referees for help with previous drafts. We are also grateful to seminar audiences at UCLA, UC San Diego, Caltech, University of Chicago, Universidad Carlos III de Madrid, Universitat Pompeu Fabra, University of Minnesota, University of Arizona, and Università Bocconi.

†Financial support from the Spanish Ministry of Education's Research Grant No. SEJ 2004-07861 while at Universidad Carlos III de Madrid is gratefully acknowledged.

*But who will monitor the monitor?* Alchian and Demsetz (1972, p. 782)

# 1 Introduction

Ann owns a restaurant. She hires Bob to tally the till every night and report back any mismatch between the till and that night's bills. Ann can motivate Bob to exert such effort and report truthfully any mismatch by sometimes secretly taking money from the till herself and offering him the following incentive scheme: if Ann took some money, she will pay Bob his wage only when he reports a mismatch; if Ann did not take any money, she will pay Bob only when a mismatch is not reported.

Bob faces a secret contract: his report-contingent wage is unknown to him a priori (it depends on whether or not Ann secretly took some money). If Bob fails to exert effort, he won't know what to report in order to secure his wage. However, if he does his job he'll discover whether or not there is a mismatch and deduce from this Ann's behavior. Only after tallying the till will Bob know what to report in order to receive his wage, which turns out to be optimally truthful.

This paper studies contracts like Bob's[1] and how they might help organizations to function productively. By allocating different information to team members, secret contracts often provide better incentives to perform with an intuitive organizational design. Thus, they give Bob incentives to acquire costly information and reveal it, and provide Ann with enough a priori knowledge to distinguish working from shirking. In general, they provide a way to "monitor the monitor" (Section 2.1), and can yield approximately efficient partnerships by appointing a "secret principal" (Section 2.2).

Consider a hypothetical organization whose individuals are subject to moral hazard but with rich communication protocols: access to (i) a disinterested mediator or machine that makes confidential, verifiable but non-binding recommendations to agents, and (ii) (linear) "money" transfers that may depend on the mediator's recommendations and individual reports (such as Bob's). A contract thus involves instructions and payments:, i.e., a way of telling people what to do and a way of rewarding them.

When can this organization actually overcome moral hazard with secret contracts?

[1] These contracts are pervasive. For instance, TSA screeners are evaluated with "covert testing" (TSA, 2004, p. 5); police use young 'drinkers' to ensure that bartenders check IDs (Cheslow, 2005).

Below, we study incentives in such a team to answer this question in various contexts. Formally, we consider contractual arrangements subject to incentive compatibility as described by Myerson's (1986) *communication equilibrium*. We obtain minimal conditions on a team's primitives—its "monitoring technology" (what Bob can see when) and individual preferences (whether Bob prefers to work or shirk)—such that incentive compatibility is not a binding constraint for the team.

Theorem 1 provides a necessary and sufficient condition on a monitoring technology—called detecting unilateral disobedience (DUD)—for every team outcome (e.g., Bob works) to be approximately enforceable, i.e., an incentive compatible outcome exists arbitrarily close to it. DUD requires that every disobedient deviation by any individual be statistically detectable with some reaction by others, although different deviations may be detected with different reactions. This key property distinguishes DUD substantively from the literature[2] (Section 3.1 has a detailed literature review). Therefore, DUD is a weak restriction. It is also generic (Theorem 2).

Secret contracts add value *not* by approximate enforcement (Corollaries 1 and 4),[3] but by allowing "monitors" to *follow* "deviators" in a hypothetical game of hide and seek, even though in fact they move simultaneously. To illustrate, suppose Bob shirks. If he also reports no mismatch then Ann can hypothetically "react" by secretly taking some money to prove him wrong, whereas if he reports a mismatch then Ann can choose not to take any money. By Theorem 1, such disobedience (e.g., Bob shirking) is detectable in this sense if and only if obedience is enforceable with secret contracts.

Restricting attention to budget-balanced transfers,[4] Theorem 3 characterizes approximate enforcement of any team outcome with a stronger condition, called identifying obedient agents (IOA). In addition to DUD, IOA requires that after any unilateral disobedience, someone can be statistically identified as obedient. IOA is weak (this is argued at the end of Section 3.2) and generic (Theorem 5), too. Intuitively, IOA provides incentives with budget balance by rewarding the innocent while punishing all others. Since IOA is necessary to deliver incentives, it exhausts the informational economies from determining "who *didn't* dunnit" rather than, say, "who dunnit."[5]

[2]For instance, conditions like *individual full rank* of Fudenberg et al. (1994) require that every deviation be detected by the same "reaction," making it more difficult to detect deviations.

[3]Even though we sometimes rely on approximation to expand contractual possibilities, our key insight is the use of *mediated transfers*—not approximating outcomes—to provide incentives.

[4]Budget balance means that the sum payments across individuals always equals zero.

[5]E.g., conditions based on *pairwise full rank* of Fudenberg et al. (1994) require that the deviator be statistically identified after every unilateral deviation, which is clearly stricter than IOA.

Theorems 6 and 8 (Section 4) extend these results in two important directions that help to clarify the differences between exact and approximate enforcement.

Theorem 6 characterizes monitoring technologies that approximately enforce a fixed outcome rather than every outcome simultaneously (Theorem 1), regardless of individual preferences. Interestingly, Theorem 6 reconciles an infinite regress inherent in monitoring. Suppose that providing incentives for a given outcome requires a monitor to detect deviations. What about the monitor's deviations? Theorem 6 answers this question by asserting that effectively the monitor's deviations are irrelevant. Indeed, if they are detectable then they can be easily discouraged with contingent payments. Otherwise, if the monitor's deviations are undetectable then the deviations themselves still detect others' deviations from the given outcome, and so they continue to fulfill the required monitoring role. Evidently, this argument also applies to the monitor's deviations from these deviations, and so forth. Theorem 6 reconciles this infinite regress by showing that under standard conditions (e.g., in a finite game) not every behavior by the monitor can have a profitable, undetectable deviation. Therefore, to approximately enforce an arbitrary outcome with infrequent monitoring, every deviation from the outcome must in principle be detectable with some monitoring behavior, but deviations away from the monitoring behavior itself need not be detectable. Heuristically, *nobody needs to monitor the monitor*.

Theorem 8 extends Theorem 6 by fixing individual preferences and finding joint conditions on preferences and the monitoring technology that characterize approximate enforcement. Intuitively, profitable deviations must be discouraged "uniformly" and "credibly." Uniform detection allows for infinitesimal deviations to be discouraged even if they are only infinitesimally detectable (Example 9). Credibility is necessary when deviations are discouraged with the use of others' actions rather than with contingent payments, and this disciplining behavior must be incentive compatible.

The paper is organized as follows. Section 2 presents two motivating examples that guide our main results. Section 3 develops the model. Section 3.1 defines DUD, characterizes its incentive properties, finds conditions for its generic satisfaction, and relates it to the literature. Section 3.2 repeats this exercise for IOA. Section 4 extends the model by characterizing exact and approximate enforcement of fixed outcomes with and without fixed preferences, and accommodates complications such as participation constraints, limited liability, and even coalitional deviations. Section 5 concludes. Omitted proofs and ancillary results appear in Appendices A and B.

# 2 Examples

We begin our analysis of secret contracts with two leading examples that attempt to capture the intuition behind our main results, Theorems 1 and 3. The first example considers an environment that typifies the strategic interaction between a principal, a worker, and a monitor. The second example suggests an intuitive way of attaining approximately efficient partnerships with budget balance.

## 2.1 Robinson and Friday

There are two agents: Robinson, who can either monitor or rest, and Friday, who can either work or shirk. A *mediating* principal makes possibly secret recommendations to the agents and enforces contingent contractual payments. Robinson (the row player) and Friday (the column player) interact according to the left bi-matrix below.

| | work | shirk |
|---|---|---|
| monitor | $2,-1$ | $-1,0$ |
| rest | $3,-1$ | $0,0$ |

Utility Payoffs

| | work | shirk |
|---|---|---|
| monitor | $1,0$ | $0,1$ |
| rest | $1/2,1/2$ | $1/2,1/2$ |

Signal Probabilities

There are two signals, $g$ and $b$, on which to condition linear transfers. Their conditional probability is given in the right bi-matrix above. In words, if Robinson monitors he observes Friday's effort, whereas if he rests then the signal is uninformative.

Although clearly the efficient profile (rest,work) is unenforceable, we can get arbitrarily close even if only Robinson observes the signal and it is not verifiable.[6] For the principal to write signal-contingent contracts, he must first solicit the realizations from Robinson, who may in principle misreport them.[7] We approximate (rest,work) by having Friday mix between working and shirking and Robinson's report-contingent payments depend on Friday's recommendation, thereby "monitoring the monitor." Specifically, the following correlated strategy is incentive compatible given $\mu \in (0,1)$:

[6] If signals are publicly verifiable, the correlated strategy $\sigma[(\text{monitor,work})]+(1-\sigma)[(\text{rest,work})]$, where $[a]$ means Dirac measure for any action profile $a$, is enforced for all $\sigma \in (0,1]$ with Holmström's (1982) *group penalties*, e.g., by paying Robinson \$2 and Friday $\$1/\sigma$ if $g$ and both agents \$0 if $b$.

[7] Now group penalties break down, since then Robinson reports $g$ and rests, hence Friday shirks. Furthermore, if Robinson was paid independently of his report then although he would happily tell the truth, he would find no reason to monitor.

(i) Robinson is told to monitor with probability $\sigma$ (and rest with probability $1-\sigma$), (ii) Friday is independently told to work with probability $\mu$ (to shirk with $1-\mu$), and (iii) the principal enforces the following *secret contract*:

| | (monitor,work) | (monitor,shirk) | (rest,work) | (rest,shirk) |
|---|---|---|---|---|
| $g$ | $1/\mu, 1/\sigma$ | $0, 0$ | $0, 0$ | $0, 0$ |
| $b$ | $0, 0$ | $1/(1-\mu), 0$ | $0, 0$ | $0, 0$ |

The table reads as follows. The leftmost column says that Robinson is paid $\$1/\mu$ if he reports $g$ and \$0 if $b$ when (monitor,work) was recommended, whereas Friday is paid $\$1/\sigma$ if $g$ is reported and \$0 if $b$, etc. Honesty and obedience to the mediator is now incentive compatible. Letting $\sigma \to 0$ and $\mu \to 1$, (rest,work) can now be approached.

Intuitively, Robinson is rewarded only when he reports $g$ if Friday was asked to work and $b$ if Friday was asked to shirk. Robinson, like Bob, faces a "trick question."

Secret contracts add value in this example because they allow different correlated strategies to detect different deviation plans, unlike just signal-contingent contracts. In other words, this is as if a correlated strategy is chosen *after* agents choose deviation plans in order to detect them. To illustrate, suppose that Robinson is asked to monitor but instead chooses to rest and report $g$. The mediator can "react" by asking Friday to shirk, which would lead to $b$ if Robinson monitored and reported truthfully. Similarly, if Robinson plans to rest and report $b$ then Friday can be asked to work instead, and Robinson's deviation is detected again.

The key idea behind Theorem 1 shows that, therefore, Robinson can be dissuaded from resting. However, with only signal-contingent contracts (Corollary 1), detecting Robinson's deviations requires the principal to fix Friday's behavior in advance. If Friday works with fixed probability $\mu$ then Robinson can rest and report $g$ with probability $\mu$. Now Robinson can deviate without being detected, and no contract contingent only on signals can induce him to monitor.

## 2.2 Secret Principal

A team has $n$ individuals. Each team member $i$ can either work ($a_i = 1$) or shirk ($a_i = 0$). Let $c > 0$ be each individual's cost of effort. Effort is not observable. Output is publicly verifiable and can be either good ($g$) or bad ($b$). The probability of $g$ equals $P(\sum_i a_i)$, where $P$ is a strictly increasing function of the sum of efforts.

Radner et al. (1986) showed that in this environment there do not exist budget-balanced output-contingent linear transfers to induce everyone to work, not even approximately. One arrangement that is not approximately efficient but nevertheless induces most people to work is appointing Holmström's principal. Call this agent 1 and define transfers as follows. For $i = 2, \ldots, n$, let $\zeta_i(g) = \overline{z}$ and $\zeta_i(b) = 0$ be agent $i$'s output-contingent linear transfer, for some $\overline{z} \geq 0$. Let agent 1's transfer equal

$$\zeta_1 \;=\; -\sum_{i=2}^{n} \zeta_i.$$

By construction, the budget is balanced. It is easy to see that everyone but agent 1 will work if $\overline{z}$ is sufficiently large. However, agent 1 has the incentive to shirk.[8]

Allowing now for secret contracts, consider the following scheme. For any small $\varepsilon > 0$, a mediator asks every individual to work (call this event $\mathbf{1}$) with probability $1 - \varepsilon$. With probability $\varepsilon/n$, he picks agent $i$ (everyone is picked with equal probability) and secretly asks him to shirk, while asking all others to work (call this event $\mathbf{1}_{-i}$). For $i = 1, \ldots, n$, let $\zeta_i(g|\mathbf{1}) = \zeta_i(b|\mathbf{1}) = 0$ be agent $i$'s contingent transfer if the mediator asked everyone to work. Otherwise, if agent $i$ was secretly told to shirk, for $j \neq i$ let $\zeta_j(g|\mathbf{1}_{-i}) = \overline{z}$ and $\zeta_j(b|\mathbf{1}_{-i}) = 0$ be agent $j$'s transfer. For agent $i$, let

$$\zeta_i = -\sum_{j \neq i} \zeta_j.$$

Clearly, this contract is budget-balanced. It is also incentive compatible. Indeed, it is clear from the contract that asking a agent to shirk is always incentive compatible. If agent $i$ is recommended to work, incentive compatibility requires that

$$\tfrac{\varepsilon}{n}(n-1)P(n-1)\overline{z} - c \;\geq\; \tfrac{\varepsilon}{n}(n-1)P(n-2)\overline{z},$$

which is satisfied if $\overline{z}$ is sufficiently large because $P$ is strictly increasing. Under this contract, everyone works with probability $1 - \varepsilon$, for any $\varepsilon > 0$, by choosing $\overline{z}$ appropriately, so everyone working is approximated with budget balanced transfers.

If a worker deviates (i.e., shirks), he lowers the probability of $g$. If a secret principal deviates (i.e., works) he raises the probability $g$. Hence, a worker's deviation changes probabilities differently from a secret principal's deviation, so after a deviation that raises the frequency of $g$, innocence can be attributed to the secret principal. Secret contracts add value by using different secret principals for different workers. This is the insight exploited by IOA to establish Theorem 3 below.

[8] This contract follows Holmström's suggestion to the letter: agent 1 is a "fixed" principal who absorbs the incentive payments to all others by "breaking" the budget constraint.

# 3 Model

This section develops the main model of secret contracts, whose purpose is to characterize a team's enforceable outcomes. Firstly, basic notation is introduced, the timing of interaction amongst team members is described explicitly, and several notions of enforcement are formally defined that will be used extensively later.

Section 3.1 then extrapolates from the leading example in Section 2.1. A notion of detection of deviation plans is introduced and the equivalence between detection and enforcement is derived in terms of a hypothetical zero-sum game of hide and seek where the hider (a deviator) moves first and the seeker (a monitor) moves second. Allowing the seeker to move second is shown to characterize the value of secret contracts. Formally, an outcome is shown to be enforceable if and only if the seeker wins in this hypothetical hide-and-seek game. A notion of "almost perfect monitoring" is also proposed, called detecting unilateral disobedience, and conditions are provided for it to obtain generically. Section 3.1 ends with a literature review.

Section 3.2 extends the results of Section 3.1 to include budget-balanced contracts in the spirit of the secret principal from Section 2.2. There, a similar hide-and-seek intuition emerges, except that now enforcement with budget balance is equated to attribution rather than just detection. Intuitively, attribution is taken to mean that it is possible to identify an obedient agent after a deviation is detected.

We begin by defining the basic strategic environment. Let $I = \{1, \dots, n\}$ be a finite set of agents, $A_i$ a finite set of actions available to any agent $i \in I$, and $A = \prod_i A_i$ the (nonempty) space of action profiles. Let $v_i(a)$ denote the utility to agent $i \in I$ from action profile $a \in A$. A *correlated strategy* is a probability measure $\sigma \in \Delta(A)$.[9]

Let $S_i$ be a finite set of *private signals* observable only by individual member $i \in I$ and $S_0$ a finite set of *publicly verifiable* signals. Let

$$S := \prod_{j=0}^{n} S_j$$

be the (nonempty) product space of all observable signals. A *monitoring technology* is a measure-valued map $\Pr : A \to \Delta(S)$, where $\Pr(s|a)$ stands for the conditional probability that $s = (s_0, s_1, \dots, s_n) \in S$ was observed given that the team played $a = (a_1, \dots, a_n) \in A$.

[9] If $X$ is a finite set, $\Delta(X) = \{\mu \in \mathbb{R}_+^X : \sum_x \mu(x) = 1\}$ is the set of probability vectors on $X$.

Assume that the team has access to *linear transfers*. An *incentive scheme* is any map $\zeta : I \times A \times S \to \mathbb{R}$ that assigns monetary transfers contingent on individuals, recommended actions, and *reported* signals. It is assumed that recommendations are verifiable.[10] Rather than focus on incentive schemes $\zeta$, we will also study *probability weighted* transfers, $\xi : I \times A \times S \to \mathbb{R}$. For any recommendation $a \in A$ with $\sigma(a) > 0$, one may think of $\xi$ as solving $\xi_i(a, s) = \sigma(a)\zeta_i(a, s)$ for some $\zeta$. For any $a \in A$ with $\sigma(a) = 0$ and $\xi(a) \neq 0$, one may think of $\xi$ as either arising from unbounded incentive schemes (i.e., $\zeta_i(a, s) = \pm\infty$) or as the limit of a sequence $\{\sigma^m\zeta^m\}$. This change of variables from $\zeta$ to $\xi$ is explained further in Section 4.1.

The timing of team members' interaction runs as follows. Firstly, agents agree upon some *contract* $(\sigma, \zeta)$ consisting of a correlated strategy $\sigma$ and an incentive scheme $\zeta$. A profile of recommendations is drawn according to $\sigma$ and made to agents confidentially and verifiably by some machine. Agents then simultaneously take some action. Taken actions are neither verifiable nor directly observable. Next, agents observe their unverifiable private signals and submit a verifiable report of their observations (given by a signal) before observing the public signal (not essential, just simplifying). Finally, recommendation- and report-contingent transfers are made according to $\zeta$.

If every agent obeys his recommendation and reports truthfully, the expected utility to agent $i$ (before recommendations are actually made) from a contract $(\sigma, \zeta)$ is

$$\sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a)\zeta_i(a, s) \Pr(s|a).$$

Of course, Mr. $i$ may disobey his recommendation $a_i$ to play some other action $b_i$ and lie about his privately observed signal. A *reporting strategy* is a map $\rho_i : S_i \to S_i$, where $\rho_i(s_i)$ is the reported signal when Mr. $i$ privately observes $s_i$. Let $R_i$ be the set of all reporting strategies for agent $i$. The *truthful reporting strategy* is the identity map $\tau_i : S_i \to S_i$ with $\tau_i(s_i) = s_i$. Thus, both $\zeta_i(a, s_{-i}, \tau_i(s_i)) = \zeta_i(a, s)$ and $\xi_i(a, s_{-i}, \tau_i(s_i)) = \xi_i(a, s)$.[11] The space of pure *deviations* for $i$ is therefore $A_i \times R_i$.

For every agent $i$ and every deviation $(b_i, \rho_i)$, the conditional probability that signal profile $s$ will be reported when everyone else is honest and plays $a_{-i} \in A_{-i}$ equals

$$\Pr(s|a_{-i}, b_i, \rho_i) \;:=\; \sum_{t_i \in \rho_i^{-1}(s_i)} \Pr(s_{-i}, t_i|a_{-i}, b_i).$$

[10] If recommendations were not directly verifiable, then agents could be asked to announce theirs as verifiable messages. However, this would involve some loss of generality (Example 2).

[11] We will often use the notation $s = (s_{-i}, s_i)$ and $a = (a_{-i}, a_i)$ for any $i$, where $s_i \in S_i$ and $s_{-i} \in S_{-i} = \prod_{j \neq i} S_j$; similarly for $A_{-i}$.

When all other agents are honest and obedient, the utility to $i$ from deviating to $(b_i, \rho_i)$ conditional on being recommended to play $a_i$ under contract $(\sigma, \zeta)$ equals

$$\sum_{a_{-i}} \frac{\sigma(a)}{\sigma(a_i)} v_i(a_{-i}, b_i) - \sum_{(a_{-i},s)} \frac{\sigma(a)}{\sigma(a_i)} \zeta_i(a, s) \Pr(s|a_{-i}, b_i, \rho_i),$$

where $\sigma(a_i) = \sum_{a_{-i}} \sigma(a) > 0$ is the probability that $a_i$ was recommended.

A team's *metering problem* is to find a contract $(\sigma, \zeta)$ that makes incentive compatible obeying recommended behavior as well as honest reporting of monitoring signals. This is captured by the following family of inequalities.

$$\forall i \in I, a_i \in A_i, (b_i, \rho_i) \in A_i \times R_i,$$
$$\sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) \;\leq\; \sum_{(a_{-i},s)} \sigma(a)\zeta_i(a, s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)). \quad (*)$$

The left-hand side reflects the *deviation gain* in terms of utility[12] for an agent $i$ from playing $b_i$ when asked to play $a_i$. The right-hand side reflects his *contractual loss* from deviating to $(b_i, \rho_i)$ relative to honesty and obedience (i.e., playing $a_i$ after being asked to do so and reporting according to $\tau_i$). Such a loss originates from two sources. On the one hand, playing $b_i$ instead of $a_i$ may change conditional probabilities over signals. On the other, reporting according to $\rho_i$ may affect conditional payments.

**Definition 1.** A correlated strategy $\sigma$ is *exactly enforceable* (or simply *enforceable*) if there exists an incentive scheme $\zeta : I \times A \times S \to \mathbb{R}$ to satisfy $(*)$ for all $(i, a_i, b_i, \rho_i)$. Call $\sigma$ *exactly enforceable with budget balance* if it is exactly enforceable and

$$\forall (a, s), \qquad \sum_{i \in I} \zeta_i(a, s) \;=\; 0.$$

A correlated strategy $\sigma$ is *approximately enforceable* if a sequence $\{\sigma^m\}$ of enforceable correlated strategies exists with $\sigma^m \to \sigma$. Call $\sigma$ *approximately enforceable with budget balance* if, in addition, every $\sigma^m$ is enforceable with budget balance.

A correlated strategy is approximately enforceable if it is the limit of exactly enforceable ones. E.g., in Section 2.1 the correlated strategy [(rest,work)] is approximately enforceable but not enforceable. Approximate enforcement with budget balance requires that the budget be balanced along the way, not just asymptotically. E.g., in Section 2.2, everybody working is approximately enforceable with budget balance, but not exactly enforceable with budget balance, even though it is exactly enforceable.

[12] Specifically, in terms of probability weighted utility, weighted by $\sigma(a_i)$. If $a_i$ is never recommended then $\sigma(a_i) = 0$ and both sides of the inequality equal zero.

## 3.1 Detection

We now provide a notion of detection that is shown to be equivalent to enforcement. A *deviation plan* for any agent $i$ is a map $\alpha_i : A_i \to \Delta(A_i \times R_i)$, where $\alpha_i(b_i, \rho_i|a_i)$ stands for the probability that $i$ deviates to $(b_i, \rho_i)$ when recommended to play $a_i$. Given $\sigma \in \Delta(A)$, let $\Pr(\sigma) \in \mathbb{R}^S$ be the vector defined by $\Pr(\sigma)(s) = \sum_a \sigma(a) \Pr(s|a)$. Intuitively, $\Pr(\sigma)$ is the vector of prior report probabilities if everyone is honest and obediently playing according to $\sigma$. Let $\Pr(\sigma, \alpha_i) \in \mathbb{R}^S$, defined pointwise by

$$\Pr(\sigma, \alpha_i)(s) \;=\; \sum_{a \in A} \sigma(a) \sum_{(b_i, \rho_i)} \Pr(s|a_{-i}, b_i, \rho_i) \alpha_i(b_i, \rho_i|a_i),$$

be the vector of prior probabilities if agent $i$ deviates from $\sigma$ according to $\alpha_i$.

A deviation plan $\alpha_i$ is *disobedient* if $\alpha_i(b_i, \rho_i|a_i) > 0$ for some $a_i \neq b_i$, i.e., it disobeys some recommendation $a_i$ with positive probability. A disobedient deviation plan may be "honest," i.e., $\rho_i$ may equal $\tau_i$ with probability one after every recommendation. A profile $\alpha = (\alpha_1, \ldots, \alpha_n)$ of deviation plans is called *disobedient* if $\alpha_i$ is disobedient for some agent $i$. Although dishonesty is arguably a form of disobedience, it will be useful in the sequel to distinguish between them.

**Definition 2** (**Detection**)**.** A deviation plan $\alpha_i$ for agent $i$ is called *undetectable* if

$$\forall \sigma \in \Delta(A), \qquad \Pr(\sigma) \;=\; \Pr(\sigma, \alpha_i).$$

Call $\alpha_i$ *detectable* if it is not undetectable, i.e., $\Pr(\sigma) \neq \Pr(\sigma, \alpha_i)$ for some $\sigma \in \Delta(A)$.

Intuitively, a deviation plan $\alpha_i$ is undetectable if the probability of reported signals induced by $\alpha_i$, $\Pr(\sigma, \alpha_i)$, coincides with that arising from honesty and obedience, $\Pr(\sigma)$, *regardless of the team's correlated strategy*, $\sigma$, assuming that others are honest and obedient. Undetectability is arguably a strong restriction on a deviation plan, making detectability a weak requirement.[13] We now give our first main definition.

**Definition 3** (**DUD**)**.** A monitoring technology Pr *detects unilateral disobedience* (DUD) if every disobedient deviation plan is detectable.

DUD is intuitively defined.[14] Formally, note that different correlated strategies may be used to decide whether or not different disobedient deviation plans are detectable.

[13] Undetectability may be defined equivalently by $\Pr(a) = \Pr(a, \alpha_i)$ for all $a \in A$ by linearity.

[14] For a slightly stronger but also mathematically more tractable version of DUD (without using reporting strategies), see Lemma B.1.

This is one important aspect that renders DUD substantially weaker than other conditions in the literature, as will soon be seen. To illustrate, consider an example.

**Example 1.** There are two publicly verifiable signals, $S = S_0 = \{x, y\}$, and two agents, $I = \{1, 2\}$. Agent 1 has two actions, $A_1 = \{U, D\}$, and agent 2 has three actions, $A_2 = \{L, M, R\}$. The conditional probability system Pr is given below.

| | $L$ | $M$ | $R$ |
|---|---|---|---|
| $U$ | 1, 0 | 0, 1 | 1/2, 1/2 |
| $D$ | 1, 0 | 0, 1 | 1/3, 2/3 |

If agent 1 plays $U$ then there is a mixed deviation by agent 2 (namely $\frac{1}{2}[L] + \frac{1}{2}[M]$, where $[\cdot]$ stands for Dirac measure) such that the conditional probability over signals equals what it would be if he played $R$. A similar phenomenon takes place when agent 1 plays $D$ (this time with the deviation $\frac{2}{3}[L] + \frac{1}{3}[M]$) or indeed regardless of agent 1's mixed strategy. It is therefore impossible to even approximately enforce $R$ with transfers contingent only on signals if agent 2 strictly prefers playing $L$ and $M$, since there always exists a profitable deviation without any contractual losses.

However, Pr detects unilateral disobedience because for any deviation plan by agent 2 there is a mixed strategy by agent 1 that detects it. By correlating agent 2's payment with agent 1's recommendation, secret contracts can keep agent 2 from knowing the proportion with which he ought to mix between $L$ and $M$ for his contractual payment to equal what he would obtain by playing $R$. It will be seen that this renders $R$ enforceable. This suggests how secret contracts can extract more information from a monitoring technology to provide incentives, even with publicly verifiable signals.

Next, we will show that DUD characterizes approximate enforcement.

**Definition 4 (PSI).** A monitoring technology Pr *provides strict incentives* (PSI) if there exists a probability weighted incentive scheme $\xi : I \times A \times S \to \mathbb{R}$ such that

$$\forall (i, a_i, b_i, \rho_i), \qquad 0 \;\leq\; \sum_{(a_{-i}, s)} \xi_i(a, s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

with a *strict inequality* whenever $a_i \neq b_i$.[15]

[15] Although no budget constraints are imposed, we could have added *expected* budget balance,

$$\sum_{(i,a,s)} \xi_i(a, s) \;=\; 0,$$

but this constraint would not bind, since adding a constant to any $\xi$ preserves its incentive properties.

By scaling $\xi$ as necessary, PSI implies that for every utility profile there is an incentive scheme so that any deviator's contractual loss outweighs his deviation gain. PSI may appear to be a rather strong condition, in contrast with the argued weakness of DUD (Example 1). As it turns out, PSI and DUD are equivalent, in fact mutually dual.

**Lemma 1.** *A monitoring technology detects unilateral disobedience if and only if it provides strict incentives.*

*Proof.* By the Alternative Theorem (Rockafellar, 1970), PSI fails if and only if there is a vector $\lambda \geq 0$ such that $\lambda_i(a_i, b_i, \rho_i) > 0$ for some $(i, a_i, b_i, \rho_i)$ with $a_i \neq b_i$ and

$$\forall (a, s), \qquad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) \;=\; 0.$$

Such a vector $\lambda$ exists if and only if the deviation plan $\alpha_i$, defined pointwise by

$$\alpha_i(b_i, \rho_i|a_i) := \begin{cases} \lambda_i(a_i, b_i, \rho_i) / \sum_{(b_i', \rho_i')} \lambda_i(a_i, b_i', \rho_i') & \text{if } \sum_{(b_i', \rho_i')} \lambda_i(a_i, b_i', \rho_i') > 0, \text{ and} \\ [(a_i, \tau_i)]\,(b_i, \rho_i) & \text{otherwise (where } [\cdot] \text{ denotes Dirac measure),} \end{cases}$$

is disobedient and undetectable: DUD fails. □

The simple proof of Lemma 1 above describes a duality between identifiability and enforceability via secret contracts. A natural corollary follows that motivates DUD from a "backward-engineering" exercise: what minimal requirement on a monitoring technology suffices to contractually overcome incentive constraints? Given $\xi$ and any completely mixed correlated strategy $\sigma \in \Delta^0(A) := \{\sigma \in \Delta(A) : \forall a \in A, \sigma(a) > 0\}$, there exists $\zeta$ with $\xi_i(a, s) = \sigma(a)\zeta_i(a, s)$ for all $(i, a, s)$. Hence, PSI is equivalent to *every* $\sigma \in \Delta^0(A)$ being (exactly) enforceable, which proves the next result.

**Theorem 1.** *A monitoring technology detects unilateral disobedience if and only if any team with any profile of utility functions can approximately enforce any correlated strategy with secret contracts.*

As Example 1 shows, DUD is not enough to provide incentives with just signal-contingent contracts, but the following strengthening is. Given a subset $B \subset A$ and an agent $i$, let $B_i := \{b_i \in A_i : \exists b_{-i} \in A_{-i} \text{ s.t. } b \in B\}$ be the projection of $B$ on $A_i$. Call a deviation plan $\alpha_i$ *$B$-disobedient* if it is disobedient at some $a_i \in B_i$. Given $\sigma \in \Delta(A)$, say Pr *detects unilateral disobedience* at $\sigma$ (DUD-$\sigma$) if $\Pr(\sigma) \neq \Pr(\sigma, \alpha_i)$ for every agent $i$ and supp $\sigma$-disobedient[16] deviation plan $\alpha_i$. Intuitively, the same $\sigma$ detects every $\alpha_i$. The proof of Theorem 1 also implies the following corollary.

[16]By definition, supp $\sigma = \{a \in A : \sigma(a) > 0\}$ is the support of $\sigma$.

**Corollary 1.** *Fix a correlated strategy $\sigma$. A monitoring technology detects unilateral disobedience at $\sigma$ if and only if any team with any profile of utility functions can enforce $\sigma$ with just "standard" signal-contingent contracts.*

Corollary 1 captures the value-added of secret contracts. By the paragraph preceding Theorem 1, DUD suffices to enforce any completely mixed $\sigma$ with secret transfers by effectively allowing the use of different $\sigma$'s to detect different $\alpha$'s, unlike standard contracts, for which the same $\sigma$ must detect every $\alpha$.[17] On the other hand, to enforce a pure-strategy profile $a$, DUD is generally not enough. Since agents receive only one recommendation under $[a]$, there is no use for secret contracts, so by Corollary 1 DUD-$[a]$ characterizes enforcement with secret as well as standard contracts.

The intermediate case where $\sigma$ has arbitrary support is discussed in Section 4.1. There, necessary and sufficient conditions are derived for exact as well as approximate enforcement. Section 4.2 extends the results further by fixing utility functions.

Genericity of DUD is established next from the number of agents' action-signal pairs. Intuitively, incentives may be provided to a given agent in three ways: (a) using only others' signals to detect his deviations (e.g., Friday), (b) using only his own reports and others' recommendations (e.g., Robinson), and (c) using both his reports and others' signals in conjunction. Theorem 2 below identifies conditions such that for every agent, at least one such way of detecting deviations is generic.

**Theorem 2.** *DUD is generic if for every agent $i$,*
(a) $|A_i| - 1 \leq |A_{-i}|\,(|S_{-i}| - 1)$ *when* $|S_i| = 1$,
(b) $|A_i|\,(|S_i| - 1) \leq |A_{-i}| - 1$ *when* $|S_{-i}| = 1$, *and*
(c) $|A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}|$ *when both* $|S_i| > 1$ *and* $|S_{-i}| > 1$.

If $|S| = 1$ then DUD is generic only if $|A| = 1$. More interestingly, DUD is generic even if $|S| = 2$, as long as agents have enough actions. Hence, a team may overcome incentive constraints (i.e., DUD, therefore Theorem 1, holds) generically even if only one individual can make substantive observations and these observations are just a binary bit of information. If others' action spaces are large enough and their actions have generic effect on the bit's probability, this uniquely informed individual may still be controlled by testing him with unpredictable combinations of others' actions.[18]

[17] Even for approximate enforcement with standard contracts the same $\sigma$ must detect all $\alpha$'s. E.g., in Example 1 there is no sequence $\{\sigma^m\}$ with $\sigma^m \to [(U, R)]$ and Pr satisfying DUD-$\sigma^m$ for all $m$.

[18] We thank an anonymous referee for urging us to emphasize this point.

We conclude this subsection by relating DUD to the literature. Broadly, DUD is an improvement in that different $\sigma$ can be used to detect different $\alpha_i$.

In a restricted setting, Legros and Matsushima (1991) and Legros and Matthews (1993) find conditions equivalent to DUD-$[a]$ (but differently interpreted) to enforce a profile $a$ with signal-contingent contracts. In repeated games, Fudenberg et al. (1994) introduced *individual full rank* (IFR). Formally, IFR (at some $\sigma$) means that for every $i$, $\Pr(\sigma) \notin \text{span}\{\Pr(\sigma, b_i, \rho_i) : (b_i, \rho_i) \neq (a_i, \tau_i)\}$, where "span" stands for linear span. Arguably, the spirit of IFR is to detect deviations away from some prescribed $\sigma$, i.e., DUD-$\sigma$.[19] IFR at $\sigma$ implies DUD-$\sigma$ but not conversely.[20] DUD is also weaker than *local* IFR (LIFR) of d'Aspremont and Gérard-Varet (1998), requiring IFR at possibly different $\sigma$ for different $i$.[21] Indeed, clearly LIFR implies DUD, and LIFR fails but DUD holds in Example 1. "Local" DUD-$\sigma$ fails there, too.

DUD is also weaker than the generalization of IFR by Kandori (2003), where agents play mixed strategies and report on the realization of such mixtures. He considers contracts contingent on those reports and signal realizations. The next example shows that secret contracts can perform strictly better even with public monitoring.

**Example 2.** One agent, three actions ($L$, $M$ and $R$), and two publicly verifiable signals ($g$ and $b$), with the following utility function and monitoring technology.

| $L$ | $M$ | $R$ |
|---|---|---|
| 0 | 2 | 0 |

Utility Payoffs

| $L$ | $M$ | $R$ |
|---|---|---|
| $1, 0$ | $1/2, 1/2$ | $0, 1$ |

Signal Probabilities

The mixed strategy $\sigma = \frac{1}{2}[L] + \frac{1}{2}[R]$ is enforceable with secret contracts but not with Kandori's contracts. Indeed, offering \$1 for $g$ if asking to play $L$ and \$1 for $b$ if asking to play $R$ makes $\sigma$ enforceable. With Kandori's contracts, the agent is asked to play $\sigma$ and then asked what he actually played before receiving any monetary rewards. The agent gains two 'utils' by playing $M$ instead and announcing that he played $L$ ($R$) if the realized signal is $g$ ($b$), with the same expected monetary payoff.[22]

[19] For instance, see Compte (1998) or Kandori and Matsushima (1998).

[20] If $|S_{-i}| < |A_i|$ for some $i$ then this holds trivially, since IFR is impossible yet DUD-$\sigma$, which requires only convex (rather than linear) independence, is possible (e.g., all the points on a circle are convexly independent). This holds even with at least as many signals as actions (e.g., consider the vectors $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$, $(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{6}, 0, \frac{1}{3}, \frac{1}{2})$ and $(\frac{1}{3}, 0, \frac{1}{6}, \frac{1}{2})$).

[21] For all $i$, LIFR uses the same correlated strategy $\sigma_i$ to detect each deviation plan $\alpha_i$ of agent $i$.

[22] Kandori's are like secret contracts if actions are *secretly* announced *before* signals are observed.

Tomala (2005) independently derives a condition comparable to DUD to prove a folk theorem. He defines detection with respect to a fixed correlated strategy using unconditional probabilities over actions and signals. He focuses on exact implementation, so for $\sigma \in \Delta^0(A)$, his version of DUD agrees broadly with ours (he proves a version of Corollary 5 below). However, he does not study approximate enforcement in general (i.e., for $\sigma \notin \Delta^0(A)$), and does not use different $\sigma$ to detect different $\alpha_i$. This issue is developed further in Theorems 6 and 8 (Section 4) below.

Finally, DUD is also *generically* weaker than the conditions cited above, i.e., it holds generically in a lower-dimensional space (see also Theorem 5 below).

## 3.2 Attribution

Let us now extend this analysis to teams with transfers subject to *budget balance*.

**Definition 5** (**Attribution**)**.** A deviation plan $\alpha_i$ for agent $i$ is *unattributable* if there exists a profile $\alpha_{-i} = (\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n)$ of deviation plans such that

$$\forall \sigma \in \Delta(A), \qquad \Pr(\sigma, \alpha_1) \;=\; \cdots \;=\; \Pr(\sigma, \alpha_i) \;=\; \cdots \;=\; \Pr(\sigma, \alpha_n).$$

Call $\alpha_i$ *attributable* if it is not unattributable, i.e., for every profile $\alpha_{-i}$ of deviation plans, there is a correlated strategy $\sigma$ and an agent $j$ such that $\Pr(\sigma, \alpha_i) \neq \Pr(\sigma, \alpha_j)$.

Intuitively, a deviation plan is unattributable if there exists a profile of opponents' deviation plans such that every unilateral deviation would lead to the same expected report probabilities. Heuristically, after an unattributable unilateral deviation, even if the fact that someone deviated is detected, anyone could have been the culprit.

**Definition 6** (**IOA**)**.** A monitoring technology Pr *identifies obedient agents* (IOA) if every disobedient deviation plan is attributable.

IOA is a stronger requirement on a monitoring technology than DUD. Indeed, DUD follows by replacing $\alpha_{-i}$ above with honesty and obedience. IOA means that any profile of disobedient deviation plans that affects the probability of reported signals must do so in a way that differs across agents. An immediate example of IOA is a team with DUD and Holmström's (1982) principal. With no actions to take or signals to observe (both $A_i$ and $S_i$ are singletons), the principal is automatically obedient. Hence, any detectable deviation can be discouraged with budget balance by rewarding him and punishing everyone else.

IOA isolates this idea and finds when the principal's role can be fulfilled internally. Theorem 3 below shows that IOA characterizes approximate enforcement with budget balance. To illustrate, recall the secret principal of Section 2.2, where if a worker shirks then good news becomes *less* likely, whereas if the secret principal works then good news becomes *more* likely. IOA holds by using *different principals for different workers*. By Theorem 3, everything is approximately enforceable with budget balance.

**Lemma 2.** *A monitoring technology identifies obedient agents if and only if it provides strict incentives with budget balance, i.e., there exists a probability weighted incentive scheme $\xi : I \times A \times S \to \mathbb{R}$ such that $\sum_i \xi_i(a,s) = 0$ for every $(a,s)$, and*

$$\forall (i, a_i, b_i, \rho_i), \qquad 0 \;\leq \sum_{(a_{-i}, s)} \xi_i(a,s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

*with a* strict inequality *whenever $a_i \neq b_i$.*

*Proof.* By the Alternative Theorem, PSI with budget balance fails if and only if $\lambda \geq 0$ and $\eta \in \mathbb{R}^{A \times S}$ exist with $\lambda_i(a_i, b_i, \rho_i) > 0$ for some $(i, a_i, b_i, \rho_i)$ with $a_i \neq b_i$ and

$$\forall (i, a, s), \qquad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) \;=\; \eta(a,s),$$

where $\eta$ is independent of $i$. Let $\Lambda = \max_{(i,a_i)} \sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) > 0$. Define

$$\alpha_i(b_i, \rho_i | a_i) := \begin{cases} \lambda_i(a_i, b_i, \rho_i)/\Lambda & \text{if } (b_i, \rho_i) \neq (a_i, \tau_i), \text{ and} \\ 1 - \sum_{(b_i,\rho_i) \neq (a_i,\tau_i)} \lambda_i(a_i, b_i, \rho_i)/\Lambda & \text{otherwise.} \end{cases}$$

By construction, $\alpha_i$ is disobedient and unattributable (using $\alpha_{-i}$): IOA fails. □

We now restrict Theorem 1 with budget balance. The proof is identical, so omitted.

**Theorem 3.** *A monitoring technology identifies obedient agents if and only if any team with any profile of utility functions can approximately enforce any correlated strategy with budget balanced secret contracts.*

Enforcement with budget-balanced standard contracts is captured by strengthening IOA as follows. Given $\sigma \in \Delta(A)$, say Pr *identifies obedient agents* at $\sigma$ (IOA-$\sigma$) if for every supp $\sigma$-disobedient deviation plan $\alpha_i$ there is a profile of deviation plans $\alpha_{-i}$ such that $\Pr(\sigma, \alpha_i) \neq \Pr(\sigma, \alpha_j)$ for some agent $j$. Intuitively, the same $\sigma$ attributes every $\alpha_i$. The next result follows easily from Theorem 3; its proof is omitted.

**Corollary 2.** *Fix a correlated strategy $\sigma$. A monitoring technology identifies obedient agents at $\sigma$ if and only if any team with any profile of utility functions can enforce $\sigma$ with signal-contingent budget-balanced contracts.*

Corollary 2 shows that secret contracts add value by allowing the use of different $\sigma$ to attribute different $\alpha_i$. The next example illustrates this point.

**Example 3.** Consider a variation on Robinson and Friday (Section 2.1) with publicly verifiable signals ($S = S_0 = \{g, b\}$) and slightly different signal probabilities:

| | work | shirk |
|---|---|---|
| monitor | $2, -1$ | $-1, 0$ |
| rest | $3, -1$ | $0, 0$ |

Utility Payoffs

| | work | shirk |
|---|---|---|
| monitor | $p, 1-p$ | $q, 1-q$ |
| rest | $1/2, 1/2$ | $1/2, 1/2$ |

Signal Probabilities

The profile (rest,work) is approximately enforceable with budget-balanced *standard* contracts if and only if $q \neq p = \frac{1}{2}$.[23] On the other hand, (rest,work) is approximately enforceable with budget-balanced *secret* contracts if and only if both $p \neq q$ and $(p - \frac{1}{2})(q - \frac{1}{2}) \geq 0$,[24] which is much weaker, and equivalent to IOA.[25]

[23] Without loss, agents mix independently. Robinson must be indifferent between monitoring and resting, so $(\frac{1}{2} - \mu p - (1-\mu)q)\Delta\zeta = 1$, where $\mu$ is the probability that Friday works, $\Delta\zeta = \zeta(g) - \zeta(b)$ and $\zeta(\omega)$ is Robinson's payment to Friday if the signal is $\omega \in \{g, b\}$. Friday will shirk unless $\sigma(p-q)\Delta\zeta = 1$ if $0 < \mu < 1$, where $\sigma$ is the probability that Robinson monitors. (If $\mu = 1$ then either $p = \frac{1}{2}$ and Robinson's constraint fails, or $p \neq \frac{1}{2}$ and $\Delta\zeta = 1/(\frac{1}{2} - p)$, so $\sigma(p-q)\Delta\zeta \to 0$ as $\sigma \to 0$ and Friday's constraint fails.) Therefore, $p \neq q$, and incentive compatibility holds as $\mu \to 1$ and $\sigma \to 0$ only if $\mu p + (1-\mu)q \to \frac{1}{2}$, i.e., $p = \frac{1}{2}$. The converse is easy.

[24] The contracts of Footnote 23 suffice if $p = \frac{1}{2}$. Otherwise (so standard contracts fail), write $\Delta\zeta$ or $\Delta\zeta'$ for the difference across signals in Robinson's payments to Friday if (monitor,work) or (monitor,shirk) was recommended, respectively. All other payments equal 0 (so rest is incentive compatible). Let $\sigma > 0$ and $\mu > 0$ mean the same as in Footnote 23. If $q > p > \frac{1}{2}$, set $\mu = 1$. Monitoring requires $1 \leq (\frac{1}{2} - p)\Delta\zeta$; working requires $1 \leq \sigma(p-q)\Delta\zeta$. Both inequalities hold if $\Delta\zeta \leq 0$ is large. If $p > q \geq \frac{1}{2}$, let $\mu < 1$. For Friday to obey we need $1 \leq \sigma(p-q)\Delta\zeta$ and $-1 \leq \sigma(q-p)\Delta\zeta'$, and for Robinson to monitor $1 \leq \mu(\frac{1}{2} - p)\Delta\zeta + (1-\mu)(\frac{1}{2} - q)\Delta\zeta'$. All inequalities hold for large $\Delta\zeta' \leq 0 \leq \Delta\zeta$. The case $p, q \leq \frac{1}{2}$ follows by symmetry; the converse is easy.

[25] Let $\alpha$ and $\beta$ be the probability that Robinson monitors if asked to rest and vice versa; $\gamma$ and $\delta$ that with which Friday works if asked to shirk and vice versa, respectively. Restricting attention to $g$ only, agents change probabilities as follows:

| Robinson | work | shirk |
|---|---|---|
| monitor | $\alpha(\frac{1}{2} - p)$ | $\alpha(\frac{1}{2} - q)$ |
| rest | $\beta(p - \frac{1}{2})$ | $\beta(q - \frac{1}{2})$ |

| Friday | work | shirk |
|---|---|---|
| monitor | $\gamma(q-p)$ | $\delta(p-q)$ |
| rest | 0 | 0 |

Clearly, LHS $\neq$ RHS for all non-zero $(\alpha, \beta, \gamma, \delta) \geq 0$ if and only if $p \neq q$ and $(p - \frac{1}{2})(q - \frac{1}{2}) \geq 0$.

To better understand IOA, let us temporarily restrict attention to *publicly verifiable monitoring technologies*, where IOA can be decomposed into DUD together with

$$\bigcap_{i \in I} C_i \;=\; \{\mathbf{0}\},$$

where $\mathbf{0}$ stands for the origin of $\mathbb{R}^{A \times S}$ and for every $i$, $C_i$ (called the *cone* of agent $i$) is the set of all vectors $\eta \in \mathbb{R}^{A \times S}$ such that for some deviation plan $\alpha_i : A_i \to \Delta(A_i)$,

$$\forall (a, s), \qquad \eta(a, s) \;=\; \sum_{b_i \in A_i} \alpha_i(b_i|a_i)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)).$$

Call this condition on $\{C_i : i \in I\}$ *non-overlapping cones* (NOC).[26] NOC means that every detectable deviation plan is attributable. Upon a detectable unilateral deviation, it may be impossible to precisely identify deviator's identity, but by NOC there must be someone to who could not have generated the statistical change. Budget-balanced incentives are provided by rewarding the obedient and punishing all others.

Decomposing IOA into DUD and NOC facilitates comparison with related literature. Fudenberg et al. (1994) impose *pairwise full rank* (PFR), implying that for *every pair* of agents, their cones do not overlap. Therefore, upon any deviation it is possible to identify the deviator's identity. On the other hand, NOC only requires that *all* agents' cones fail to overlap simultaneously. Thus, it is possible that two agents' cones overlap, i.e., their intersection is larger than just the origin, and violate PFR but still provide incentives with budget balance. In general, NOC does not even require that there always be two agents whose cones fail to overlap, in contrast with *local compatibility* of d'Aspremont and Gérard-Varet (1998), as Figure 1 below illustrates.

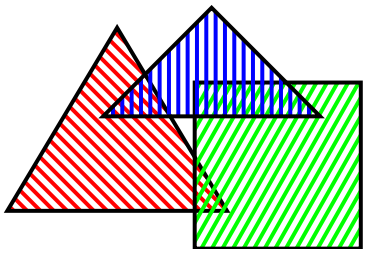

Figure 1: A cross-section of three non-overlapping cones in $\mathbb{R}^3$ (pointed at the origin behind the page) such that every pair of cones overlaps.

[26] To see that IOA is equivalent to DUD plus NOC, notice firstly that IOA clearly implies DUD. By IOA, if a deviation plan $\alpha_i$ is unattributable then it is obedient, hence undetectable, and NOC follows. Conversely, NOC implies that every unattributable $\alpha_i$ is undetectable. By DUD every undetectable $\alpha_i$ is obedient. Hence, every unattributable $\alpha_i$ is obedient and IOA follows.

NOC can be translated to an equivalent condition with dual economic interpretation that provides useful insights into the contractual role played by NOC, as shown next.

**Definition 7.** A verifiable monitoring technology Pr *clears every budget* (CEB) if given $K : A \times S \to \mathbb{R}$ there exists $\xi : I \times A \times S \to \mathbb{R}$ such that

$$\begin{aligned} \forall (a,s), \qquad & \sum_{i \in I} \xi_i(a,s) \;=\; K(a,s), \quad \text{and} \\ \forall (i,a_i,b_i), \qquad & 0 \;\leq \sum_{(a_{-i},s)} \xi_i(a,s)(\Pr(s|b_i,a_{-i}) - \Pr(s|a)). \end{aligned}$$

The function $K(a,s)$ may be regarded as a budgetary surplus or deficit for each combination of recommended action and realized signal. Intuitively, CEB means that any budget can be attained by some payment scheme that avoids disrupting any incentive compatibility constraints. As it turns out, this is equivalent to NOC.

**Corollary 3.** *A publicly verifiable monitoring technology has non-overlapping cones if and only if it clears every budget.*

This result helps clarify the role of DUD and NOC in Theorem 3. By Theorem 1, DUD characterizes approximate enforcement of any correlated strategy $\sigma$. However, the team's budget may not be balanced ex post. NOC guarantees the existence of a further contract to absorb any budgetary deficit or surplus of the original contract without violating any incentive constraints. Therefore, the original contract plus this further contract can now approximately enforce $\sigma$ with ex post budget balance.[27]

Without verifiability, a decomposition of IOA into two separate parts does not emerge naturally. Indeed, it is not difficult to see that NOC plus DUD is sufficient but not necessary for IOA. Necessity fails in general because there may exist dishonest but otherwise obedient deviations that do not directly affect anyone's utility, and as such IOA allows them to remain unattributable even if detectable. With verifiability, every deviation may in principle affect agents directly. To illustrate, consider an example.

**Example 4.** Take a team with three agents such that $A_i$ is a singleton for every agent $i$, so IOA is automatically satisfied. There are no public signals and each agent observes a binary private signal: $S_i = \{0,1\}$ for all $i$. The monitoring technology is

$$\Pr(s) := \begin{cases} \frac{6}{25} & \text{if } \sum s_i = 3 \\ \frac{3}{25} & \text{if } \sum s_i = 1 \text{ or } 2 \\ \frac{1}{25} & \text{if } \sum s_i = 0 \end{cases}$$

[27] A similar argument is provided by d'Aspremont et al. (2004) for Bayesian mechanisms.

The following is a profile of (trivially obedient) unattributable deviation plans that are also detectable, violating NOC. Suppose that agent $i$ deviates by lying with probability $2/5$ after observing $s_i = 1$ and lying with probability $3/5$ after observing $s_i = 0$. For every agent $i$, the joint distribution of reported private signals becomes:

$$\Pr(s) = \begin{cases} \frac{27}{125} & \text{if } \sum s_i = 3 \\ \frac{18}{125} & \text{if } \sum s_i = 2 \\ \frac{12}{125} & \text{if } \sum s_i = 1 \\ \frac{8}{125} & \text{if } \sum s_i = 0 \end{cases}$$

Genericity of IOA is discussed next. To motivate, consider the following example.

**Example 5.** Change Example 3 such that only Robinson observes the signal. Now it is impossible to approximately enforce (rest,work) with budget balance.[28]

Hence, IOA fails. Unfortunately, it gets worse: the example is not pathological.

**Theorem 4.** *Identifying obedient agents is impossible with only two agents and no public information.*

*Proof.* Fix an arbitrary action profile $\widehat{a} \in A$ and consider the following disobedient deviation plan $\alpha_i$ for every agent $i$: always play $\widehat{a}_i$ regardless of the mediator's recommendation $a_i$ and report $s_i$ with probability $\Pr(s_i|a_i, \widehat{a}_{-i}) = \sum_{s_{-i}} \Pr(s|a_i, \widehat{a}_{-i})$ independently of the actual signal realization. If any agent $i$ unilaterally deviates according to $\alpha_i$, the probability of reported signals becomes

$$\Pr(s|a, \alpha_i) = \begin{cases} \Pr(s_1|\widehat{a}) \Pr(s_2|\widehat{a}) & \text{if } a_1 = \widehat{a}_1 \text{ and } a_2 = \widehat{a}_2 \\ \Pr(s_1|\widehat{a}) \Pr(s_2|\widehat{a}_1, a_2) & \text{if } a_1 = \widehat{a}_1 \text{ and } a_2 \neq \widehat{a}_2 \\ \Pr(s_1|a_1, \widehat{a}_2) \Pr(s_2|\widehat{a}) & \text{if } a_1 \neq \widehat{a}_1 \text{ and } a_2 = \widehat{a}_2 \\ \Pr(s_1|a_1, \widehat{a}_2) \Pr(s_2|\widehat{a}_1, a_2) & \text{if } a_1 \neq \widehat{a}_1 \text{ and } a_2 \neq \widehat{a}_2 \end{cases}$$

These probabilities are the same regardless of who deviates, hence IOA fails. □

Theorem 4 simply says that with two agents and no public signals it is always possible to blame the other agent for a deviation. Since it is impossible to identify who deviated, by elimination it is also impossible to identify who did not deviate.

[28] Just as in Footnote 25, for Friday to work we require that $1 \leq \sigma(p - q)\Delta\zeta$. For Robinson to monitor, we require that Friday mixes between working and shirking, so $0 < \mu < 1$. Robinson's incentive constraints when asked to monitor so that he exerts the effort and reports truthfully are $1 \leq \mu(\alpha - p)\Delta\zeta + (1 - \mu)(\alpha - q)\Delta\zeta'$ for every $\alpha \in [0, 1]$. Here $\alpha$ represents Robinson's ability to lie after resting. Choosing $\alpha = q$ yields $1 \leq \mu(q - p)\Delta\zeta$, but this is inconsistent with $1 \leq \sigma(p - q)\Delta\zeta$.

Fortunately, IOA is almost always satisfied beyond this restricted environment, as the next result shows. Reorder the set $I$ of agents so that $i < j$ if $|S_i| \leq |S_j|$. Let $K = \{1, 2, \ldots, k\}$ be the subset of agents with $|S_i| = 1$, i.e., those who do not observe a private signal. Finally, reorder $K$ if necessary so that $i < j$ if $|A_i| \leq |A_j|$.

**Theorem 5.** *IOA is generic if the conditions for Theorem 2 are satisfied and*

$$\sum_{i=1}^{n}(|A_i|\,|S_i|)^2 - 1 - \chi_n(|A_n|^2\,|S_n| - 1) \;\leq\; (n-1)\,|A|\,|S| - (k-1)\,|A| + |A_k| \sum_{i=1}^{k-1} |A_i|\,,$$

*where $\chi_n = 1$ if $|S_{-n}| = 1$ and 0 otherwise, and agents are ordered as above.*

It is not difficult to see that if agent 1 is a principal, i.e., $|A_1|\,|S_1| = 1$, then IOA is generic if and only if DUD is generic. Intuitively, Theorem 5 holds when actions and signals are allocated relatively evenly across agents. This condition is weaker than others in the literature. To help understand the result, consider some examples.

**Example 6.** If every agent has the same number of actions, so $|A_i| = m$ for all $i$, and all available signals are publicly verifiable information, so $|S| = |S_0| = \ell$, then IOA is generic when $nm^2 - 1 \leq (n-1)\ell m^n - (n-1)m^n + (n-1)m^2$, or equivalently, $m^2 - 1 \leq (n-1)(\ell - 1)m^n$, which holds for all $\ell, n > 1$ and $m \geq 1$. Hence, IOA is generic *even with just two agents and two public signals*. Now suppose instead that $|S| = |S_n| = \ell > 1$, so only agent $n$ observes the signals. By Theorem 2, DUD is generic if $m(\ell - 1) \leq m^{n-1} - 1$, which holds for all $m > 1$ if $\ell = 2 < n$. The second condition for IOA simplifies to $m^2\ell^2 \leq m^n((\ell - 1)(n - 1) + 1)$, which also holds for all $m > 1$ if $\ell = 2 < n$. Therefore, IOA is generic *even if only one agent observes a binary private signal* as long as there are at least three agents.

We end this section with a discussion of some literature related to IOA. Broadly, IOA improves on previous results by (i) using different strategies to detect/attribute different deviations and (ii) not requiring a deviator's identity. Specifically, the relevant literature is divided into contract theory, mechanism design and repeated games.

In contract theory, Legros and Matsushima (1991) characterize exact enforcement with budget-balanced standard contracts and verifiable signals, but their condition is difficult to interpret, and one in terms of attribution is not suggested. Legros and Matthews (1993) study approximate enforcement with standard budget-balanced contracts and deterministic output, but fail to provide necessary conditions, and again do not discuss attribution or rewarding the innocent. Finally, d'Aspremont

and Gérard-Varet (1998) find stronger conditions (Figure 1) in a more restricted setting, using linear methods. IOA subsumes these contributions in a more general environment. IOA is also generic in a lower-dimensional space.

In mechanism design, d'Aspremont et al. (2004) provided necessary and sufficient conditions for budget-balanced implementation. Some of the results here have a similar flavor, such as Corollary 3. Independently from this paper, Kosenok and Severinov (forthcoming) extend d'Aspremont et al. (2004) to include individual rationality, and propose a condition they call weak identifiability (WI) in the context of mechanism design. Intuitively, WI implies that no profile of unilateral deviations (from truthful reporting strategies) can change the distribution of reports in the same way.

There are important differences between their work and ours, partly due to the difference between moral hazard and adverse selection. Firstly, we consider outcomes that are not necessarily pure-strategy profiles. This permits the use of different outcomes to attribute different deviations, unlike the mechanism design literature, which just enforces honest reporting. For IOA to fail, there must be a disobedient deviation that is unattributable across *all* recommended action profiles, so IOA can be satisfied even if WI is violated at a given action profile. Secondly, this paper studies both exact and approximate enforcement, whereas the mechanism design literature only considers exact enforcement. Thirdly, WI requires attribution with respect to *every* deviation, in contrast to IOA, which requires attribution only with respect to disobedient ones. Therefore, since agents' signals are not necessarily verifiable in this paper, IOA is not generally bound by the decomposition of Corollary 3, unlike the mechanism design literature. Finally, the "types" in this paper are endogenous.

As for repeated games, IOA was compared to PFR just before Figure 1, so IOA subsumes Fudenberg et al. (1994), Kandori and Matsushima (1998), Kandori (2003) and Kandori and Obara (2006). Independently, Aoyagi (2005) and Tomala (2005) use stronger conditions than IOA to prove a folk theorem for repeated games with private monitoring and mediated communication. Aoyagi's dynamic strategies rely on "$\varepsilon$-perfect" monitoring, and fail if monitoring is costly, one-sided, or public, for instance. Tomala considers a class of recursive equilibria that render his problem similar to ours by interpreting patient agents' continuation payoffs as linear transfers. His folk theorem requires budget balance for every weighted sum of transfers, which makes his condition much stronger, comparable to PFR. In addition, he assumes that every deviation is attributable, whereas IOA only attributes disobedient ones.

# 4 Discussion

This section makes four comments. Firstly, the previous section's results are extended to correlated strategies with restricted support. Secondly, these results are further extended by restricting attention to a fixed profile of utility functions. Thirdly, we comment on individual rationality and limited liability. Finally, we discuss collusion and characterize contracts that dissuade multilateral deviations.

## 4.1 Exact versus Approximate Enforcement

Next, we characterize exact enforcement of a fixed correlated strategy for any utility profile. Fix two sets of action profiles $B, C \subset A$. A deviation plan $\alpha_i$ is *$C$-detectable* if $\Pr(\sigma) \neq \Pr(\sigma, \alpha_i)$ for some $\sigma \in \Delta(A)$ with supp $\sigma \subset C$. Say Pr *$C$-detects unilateral $B$-disobedience* ($\text{DUD}_B^C$) if every $B$-disobedient deviation plan is $C$-detectable. (We will call $A$-detection simply detection, and write $\text{DUD}_B^A$ as $\text{DUD}_B$.) For instance, $\text{DUD}_A^A$ is just DUD, and $\text{DUD}_{\{a\}}^{\{a\}}$ equals DUD-$[a]$. Consider another example.

**Example 7.** There are two agents and two publicly verifiable signals, with the monitoring technology below. (It is Example 1 with an added row.)

| | $L$ | $M'$ | $R$ |
|---|---|---|---|
| $U$ | 1, 0 | 0, 1 | 2/3, 1/3 |
| $M$ | 1, 0 | 0, 1 | 1/2, 1/2 |
| $D$ | 1, 0 | 0, 1 | 1/3, 2/3 |

Let $A = \{U, M, D\} \times \{L, M', R\}$, $B = \{U, M\} \times \{L, M', R\}$, $C = \{U, D\} \times \{L, M', R\}$. Clearly, $\text{DUD}_A^A$ fails here, since $\frac{1}{2}[U] + \frac{1}{2}[D]$ is statistically indistinguishable from $M$. Also, $\text{DUD}_B^A$ fails because a plan to play $\frac{1}{2}[U] + \frac{1}{2}[D]$ when asked to play $M$ is $A$-undetectable and $B$-disobedient. However, $\text{DUD}_C^C$ does hold, since there is no undetectable deviation from $D$ or $U$ by the row player. (DUD-$\sigma$ fails for every $\sigma$.)

**Corollary 4.** *Fix any subset $B \subset A$. A monitoring technology $B$-detects unilateral $B$-disobedience if and only if any team with any profile of utility functions can exactly enforce every (if and only if some fixed) correlated strategy with support equal to $B$.*

Therefore, every correlated strategy with support equal to a subset of $B$ is automatically approximately enforceable, just as with Theorem 1. By Corollary 4, *only the support of a correlated strategy matters* for its enforcement regardless of preferences.

Having characterized exact enforcement, we proceed with approximate enforcement. By Corollary 4, existence of some $C \supset B$ such that $\text{DUD}_C^C$ holds clearly yields a sufficient condition. However, this is unnecessary. To motivate, consider an example.

**Example 8.** Two agents, two public signals, the following monitoring technology:

| | $L$ | $M$ | $R$ |
|---|---|---|---|
| $U$ | 1, 0 | 1, 0 | 1, 0 |
| $D$ | 1, 0 | 0, 1 | 0, 1 |

The action profile $(U, L)$ is not enforceable for every utility profile, since $\text{DUD}_{\{(U,L)\}}^{\{(U,L)\}}$ clearly fails. Indeed, playing $D$ when asked to play $U$ is $\{(U, L)\}$-undetectable. It is also easy to see that $\text{DUD}_C^C$ fails, too, for every $C \supset \{(U, L)\}$. However, $(U, L)$ is approximately enforceable for every utility profile, since either $[(D, M)]$ or $[(D, R)]$ can be used to detect $\{(U, L)\}$-disobedient deviations. No contract can induce the row player to choose $M$ if $R$ is strictly dominant, say, but this is unimportant as long as the row player chooses either $M$ or $R$ when asked to do so. The key condition satisfied here is that every $\{(U, L)\}$-disobedient deviation plan is detectable ($\text{DUD}_{\{(U,L)\}}$).

In general, it should be clear that $\text{DUD}_B$ is necessary for approximate enforcement, but perhaps it is not so clear that it is also sufficient, as the next result shows.

**Theorem 6.** *Fix any subset $B \subset A$. A monitoring technology detects unilateral $B$-disobedience if and only if any team with any profile of utility functions can approximately enforce every (if and only if some fixed) correlated strategy with support in (equal to) $B$ with secret contracts.*

Hence, still only the support of a correlated strategy matters for approximate enforcement regardless of preferences. Clearly, Theorem 1 is a special case of Theorem 6 (as well as Corollary 4) when $B = A$. Example 8 illustrates the insight behind Theorem 6 and gives intuition for its proof. Suppose that, to detect deviations from $a_i \in B_i$, some $a_j \notin B_j$ is played infrequently by $j \neq i$. What if $a_j$ itself has a profitable, undetectable deviation $\alpha_j(a_j) \in \Delta(A_j)$? After all, $\text{DUD}_B$ says nothing about detection outside $B$. If such $\alpha_j(a_j)$ exists, playing it instead of $a_j$ *still detects deviations* from $a_i$ by virtue of being undetectable. Similarly, undetectable deviations from $\alpha_j(a_j)$ detect deviations from $a_i$, and so on. Proceeding iteratively, since the game is finite there must be detecting behavior without a profitable, undetectable deviation.

We end this subsection by remarking that Corollary 4 and Theorem 6 generalize easily with appropriate versions of IOA after detection is replaced with attribution.

## 4.2 Fixed Utility Functions

Throughout this section, let us fix a profile $v : I \times A \to \mathbb{R}$ of utility functions. A natural weakening of the previous results might be to allow for undetectable deviation plans as long as they are *unprofitable*. Exact enforcement amounts to the following.

**Corollary 5.** *A correlated strategy $\sigma$ is enforceable with secret contracts if and only if every supp $\sigma$-undetectable deviation plan $\alpha_i$ of any agent $i$ is $\sigma$-unprofitable, i.e.,*

$$\Delta v_i(\sigma, \alpha_i) \;:=\; \sum_{(a,b_i,\rho_i)} \sigma(a)\alpha_i(b_i,\rho_i|a_i)(v_i(a_{-i},b_i) - v_i(a)) \;\leq\; 0.$$

The proof of this claim is comparable to previous ones, therefore omitted. Given an enforceable correlated strategy $\sigma$, we now ask how large transfers must be to enforce it. To this end, let us introduce some notation. Let $\mathscr{D}_i = \Delta(A_i \times R_i)^{A_i}$ be the space of deviation plans $\alpha_i$ for a agent $i$ and $\mathscr{D} = \prod_i \mathscr{D}_i$ be the set of profiles of deviation plans $\alpha = (\alpha_1, \ldots, \alpha_n)$. For any deviation plan $\alpha_i \in \mathscr{D}_i$ and any $\sigma \in \Delta(A)$, write

$$\|\Delta \Pr(\sigma, \alpha_i)\| \;:=\; \sum_{s\in S} \Big| \sum_{(a,b_i,\rho_i)} \sigma(a)(\alpha_i(b_i,\rho_i|a_i) \Pr(s|a_{-i},b_i,\rho_i) - \Pr(s|a)) \Big|.$$

This norm summarizes the difference in signal probabilities between abiding by $\sigma$ and deviating to $\alpha_i$. A correlated strategy $\sigma$ is called *enforceable within* some vector $z \in \mathbb{R}^I_+$ if there exists a scheme $\xi : I \times A \times S \to \mathbb{R}$ to satisfy $(*)$ and

$$\forall (i,a,s), \qquad -\sigma(a)z_i \leq \xi_i(a,s) \leq \sigma(a)z_i.$$

Next, we provide a lower bound on $z$ so that $\sigma$ is enforceable within $z$.

**Theorem 7.** *(i) A correlated strategy $\sigma$ is enforceable within $z \in \mathbb{R}^I_+$ if and only if*

$$V_\sigma(z) \;:=\; \max_{\alpha\in\mathscr{D}} \sum_{i\in I} \Delta v_i(\sigma,\alpha_i) - \sum_{(i,a)} z_i\sigma(a)\,\|\Delta \Pr(a,\alpha_i)\| \;=\; 0.$$

*(ii) If $\sigma$ is enforceable then $V_\sigma(z) = 0$ for some $z \in \mathbb{R}^I_+$. If not then $\sup_z V_\sigma(z) > 0$.*
*(iii) A correlated strategy $\sigma$ is enforceable if and only if $\overline{z}_i < +\infty$ for every $i$, where*

$$\overline{z}_i := \sup_{\alpha_i\in\mathscr{F}_i} \frac{\max\{\Delta v_i(\sigma,\alpha_i), 0\}}{\sum_a \sigma(a)\,\|\Delta \Pr(a,\alpha_i)\|} \quad \text{if } \mathscr{F}_i := \{\alpha_i : \textstyle\sum_a \sigma(a)\,\|\Delta \Pr(a,\alpha_i)\| > 0\} \neq \emptyset$$

*and, whenever $\mathscr{F}_i = \emptyset$, $\overline{z}_i := +\infty$ exactly when $\max_{\alpha_i} \Delta v_i(\sigma,\alpha_i) > 0$.*[29]
*(iv) If $\overline{z}_i < +\infty$ for every $i$ then $V_\sigma(z) = 0$ if and only if $z_i \geq \overline{z}_i$ for all $i$.*

[29] Intuitively, $\mathscr{F}_i$ is the set of all supp $\sigma$-detectable deviation plans available to agent $i$.

Theorem 7 quantifies the wedge that transfers require to enforce a given correlated strategy with punishments and rewards. It implies that supp $\sigma$-detectability, hence also enforceability by Corollary 5, is captured by $\sum_a \sigma(a) \|\Delta \Pr(a, \alpha_i)\| > 0$. By Corollary 1, enforcement with signal-contingent transfers is captured by the stronger detectability condition that $0 < \|\Delta \Pr(\sigma, \alpha_i)\| \leq \sum_a \sigma(a) \|\Delta \Pr(a, \alpha_i)\|$. Hence, a version of Theorem 7 holds with signal-contingent transfers and $\|\Delta \Pr(\sigma, \alpha_i)\|$ instead of $\sum_a \sigma(a) \|\Delta \Pr(a, \alpha_i)\|$. Finally, Theorem 7 (*iii*) clearly implies the following.

**Corollary 6.** *Each supp $\sigma$-undetectable deviation plan is $\sigma$-unprofitable if and only if there exists $z \geq 0$ such that $\Delta v_i(\sigma, \alpha_i) \leq z \sum_a \sigma(a) \|\Delta \Pr(a, \alpha_i)\|$ for all $i$ and $\alpha_i$, that is, utility gains from any deviation are uniformly outweighed by monetary losses.*

Next, we characterize approximate enforcement of a correlated strategy for a fixed profile of utility functions. Corollary 5 might suggest that a correlated strategy $\sigma$ is enforceable if every $\sigma$-profitable deviation plan is detectable. However, the next example shows that approximate enforcement imposes a stronger requirement.

**Example 9.** Consider a variation on Robinson and Friday (Section 2.1):

| | work | shirk | solitaire |
|---|---|---|---|
| monitor | $0, -1$ | $0, 0$ | $0, 0$ |
| rest | $0, -1$ | $0, 0$ | $0, -1$ |

Utility Payoffs

| | work | shirk | solitaire |
|---|---|---|---|
| monitor | $1, 0$ | $0, 1$ | $1, 0$ |
| rest | $1/2, 1/2$ | $1/2, 1/2$ | $1/2, 1/2$ |

Signal Probabilities

Assume for simplicity that the signal is *publicly verifiable* and Robinson's utility is constant. Clearly, the profile (rest,work) is not enforceable because a deviation by Friday to shirk is [(rest,work)]-profitable and {(rest,work)}-undetectable. Moreover, (rest,work) is *not approximately enforceable* either. Indeed, for Friday to ever work Robinson must monitor with positive probability. But then no contract can discourage Friday from playing solitaire instead of working, since playing solitaire when asked to work is undetectable and weakly dominant. On the other hand, every [(rest,work)]-profitable deviation plan is detectable.

Removing solitaire from Example 9 restores approximate enforcement of (rest,work). This occurs not because every (rest,work)-profitable deviation is detectable (it is true with or without solitaire), but because it is *uniformly detectable*, i.e., the utility gains from every (rest,work)-profitable deviation by Friday are uniformly outweighed by monetary losses when Robinson monitors, in line with Corollary 6. The next result characterizes approximate enforcement with "uniform, credible" detection.

**Theorem 8.** *A correlated strategy $\sigma$ is approximately enforceable if and only if there exists $z \geq 0$ such that every $\sigma$-profitable deviation plan $\alpha_i$ is detectable by some correlated strategy $\mu$ for which both*
(i) $\Delta v_i(\mu, \alpha_i) < z \sum_a \mu(a) \|\Delta \Pr(a, \alpha_i)\|$ *and*
(ii) $\Delta v_j(\mu, \alpha_j) \leq z \sum_a \mu(a) \|\Delta \Pr(a, \alpha_j)\|$ *for every other agent $j$ and deviation $\alpha_j$.*

Intuitively, Theorem 8 says that to approximately enforce a correlated strategy, it is necessary and sufficient that all its profitable deviations be discouraged both (i) uniformly and (ii) credibly. As before, different behavior may be used to detect different deviations by an agent.[30] Formally, uniform detection means that for the same fixed $z$, every deviation plan $\alpha_i$ must impact the magnitude of $z$-weighted probabilistic changes enough to outweigh its deviation gains. Therefore, transfers bounded within $z$ can provide incentives against all $\sigma$-profitable deviations, perhaps with different $\mu$ for different $\alpha_i$.

To explain the need for credibility, compare this result with Theorem 6, where "credible monitoring" is unnecessary. There, *every* disobedient deviation is potentially profitable, so ought to be detectable. Here, with fixed utility functions, even if some disobedient deviation plan $\alpha_i$ is undetectable, it may nonetheless be discouraged with behavior $\mu$ by others that makes the deviation unprofitable (as in a correlated equilibrium without transfers). However, if this specific behavior is not credible then there may exist a $\mu$-profitable deviation plan $\alpha_j$ by some other agent such that $\alpha_i$ becomes profitable once again given $\mu$ and $\alpha_j$.

We end this subsection by noting without proof that all previous results hold also with budget balance using the same arguments, replacing detection with attribution and $\sum_a \sigma(a) \|\Delta \Pr(a, \alpha_i)\|$ with $\min_\eta \sum_{(i,a)} \sigma(a) \|\Delta \Pr(a, \alpha_i) - \eta(a)\|$, where $\eta \in \mathbb{R}^{A \times S}$. Now, instead of the change in transfers to an agent after a deviation being bounded by the magnitude of the change in the probability over signals, it is bounded by the residuals of a least-absolute-deviations regression of the probability changes on $A$. This amount is clearly smaller than the magnitude of the dependent variable, i.e., the probability change. Intuitively, budget balance implies that only the variation across agents of the deviations' effect on signals can be used to provide incentives to discourage them, rather than the deviations' effects themselves.

[30] To see that credibility matters, simply add a row to the table in Example 8 above with utility payoffs $\boxed{-1,-1 \mid -1,0 \mid -1,-1}$ and signal probabilities $\boxed{1,0 \mid 0,1 \mid 1,0}$. Now there is an action for Robinson that is strictly dominated and indistinguishable from monitoring, yet uniformly detects all of Friday's (rest,work)-profitable deviations.

## 4.3 Participation and Liability

Individual rationality is also amenable to our study of incentives, and captured by the following family of linear inequalities:

$$\forall i \in I, \qquad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \xi_i(a,s) \Pr(s|a) \;\geq\; 0.$$

Without budget balance, since agents can be paid lump sums to become indifferent between belonging to the team and forsaking it, individual rationality constraints do not bind. Hence, suppose the team's budget must be balanced ex post.

**Theorem 9.** *Consider a team subject to incentive compatibility and budget balance. Participation is not a binding constraint if $\sum_i v_i(a) \geq 0$ for all $a$.*

This theorem generalizes standard results (e.g., d'Aspremont and Gérard-Varet, 1998, Lemma 1) to our setting. Next, we characterize enforcement subject to participation, liability and budget constraints. The proof is similar to previous ones, hence omitted.

**Theorem 10.** *The correlated strategy $\sigma$ is enforceable with budget balance and individual rationality within $z$ if and only if for every $\alpha \in \mathscr{D}$, $\pi \in \mathbb{R}^n_+$ and $\eta \in \mathbb{R}^{A \times S}$,*

$$\sum_{i \in I} \Delta v_i(\sigma, \alpha_i) \;\leq\; \sum_{i \in I} \pi_i v_i(\sigma) + \sum_{(i,a)} z_i \sigma(a) \left\| \Delta \Pr(a, \alpha_i) - \eta(a) - \pi_i \Pr(a) \right\|.$$

To interpret Theorem 10, consider a group of agents within which deviations are internally unattributable. This result says that the deviations gains for agents in the group must be compensated by the liability of agents outside the group. Indeed, let $\alpha \in \mathscr{D}$ be such that $\Delta \Pr(a, \alpha_i)$ coincides for all $a$ and $i$ in some subset $t \subsetneq I$. Let $\pi_i = \pi$ for $i \in t$ and choose $\eta$ so that $\| \Delta \Pr(a, \alpha_i) - \eta(a) - \pi \Pr(a) \| = 0$ for all $a$ and $i \in t$. If $z_j = 0$ for all $j \notin t$ then the above inequality fails if $\sigma$ is not a correlated equilibrium. Hence, $z_j$ must be positive for some $j \notin t$, i.e., the deviation gains from $\alpha$ are compensated by the liability of agents *outside* $t$.

Next, we study one-sided limited liability given $z \in \mathbb{R}^I_+$, by considering constraints of the form $\xi_i(a,s) \leq \sigma(a) z_i$. A team's *total liability* is defined by $\widehat{z} = \sum_i z_i$. *Without participation constraints*, Theorem 5 of Legros and Matsushima (1991) and Theorem 4 of Legros and Matthews (1993) easily generalize to this setting.

**Theorem 11.** *In the absence of participation constraints, only total (one-sided) liability affects a team's enforceable outcomes, not the distribution of liability.*

Including participation constraints leads to the following characterization.

**Theorem 12.** *The correlated strategy $\sigma$ is enforceable with budget balance, individual rationality and one-sided liability limited by $z$ if and only if*

$$\sum_{i\in I}\Delta v_i(a,\alpha_i) \;\leq\; \sum_{i\in I}\pi_i(v_i(\sigma)-z_i)+\widehat{\eta}\sum_{i\in I}z_i$$

*for every $(\alpha,\pi,\widehat{\eta})$ such that $\alpha$ is a profile of deviation plans, $\pi=(\pi_1,\ldots,\pi_n)\geq 0$, and $\widehat{\eta}:=\sum_s \max_i\{\pi_i\Pr(s|a)-\Delta\Pr(s|a,\alpha_i)\}$.*

A similar intuition as in Theorem 10 regarding liability applies here, too. If a subset $t$ of individuals can deviate indistinguishably then others must accept liability for it. Theorem 10 also generalizes Theorems 9 and 11, as the next result shows.

**Corollary 7.** *Suppose that $\sigma$ is enforceable with budget balance, individual rationality and one-sided liability limited by $z$.* (i) *If $v_i(\sigma)\geq z_i$ then agent $i$'s participation is not a binding constraint.* (ii) *The distribution of liability does not matter within the subset $t$ of agents whose participation constraint is not binding, i.e., $\sigma$ is also enforceable with budget balance, individual rationality and one-sided liability limited by any $z'$ with $z_j=z'_j$ for $j\in I\setminus t$ and $\sum_{i\in t}z_i=\sum_{i\in t}z'_i$.*

## 4.4 Coalitional Deviations

A notable weakness of secret contracts is not being collusion-proof. To illustrate, in our leading example (Section 2.1) Robinson and Friday could communicate "extra-contractually" to break down the incentives that secrets tried to provide.[31] On the other hand, collusion is a problem for contracts in general. For instance, the scheme proposed by Cremer and McLean (1988) is not collusion-proof for similar reasons.

To study collusion-proof contracts, assumptions must be made regarding coalitions' contractual ability. We will assume that every coalition $t$ maximizes some given *coalitional utility function* $v_t:A\to\mathbb{R}$, quasilinear in monetary transfers.[32]

[31] The following incentive scheme deters such communication between Robinson and Friday (Friday prefers misreporting his signal to Robinson) while approximately enforcing (rest,work).

| | (monitor,work) | (monitor,shirk) | (rest,work) | (rest,shirk) |
|---|---|---|---|---|
| $g$ | $1/\mu, 1/\sigma$ | $0, 1/\sigma$ | $1/2\mu, 0$ | $0, 1/2(1-\sigma)$ |
| $b$ | $0, 0$ | $1/(1-\mu), 0$ | $0, 1/(1-\sigma)$ | $1/2(1-\mu), 1/2(1-\sigma)$ |

[32] This assumption is standard. See for instance, Che and Kim (2006) and references therein. The purpose of this section is not to derive a meaningful utility for coalitions, but to use one.

**Definition 8.** A correlated strategy $\sigma$ is *strongly enforceable* if there is an incentive scheme $\zeta : I \times A \times S \to \mathbb{R}$ such that

$$\forall t \subset I, a_t \in A_t, (b_t, \rho_t) \in A_t \times R_t,$$
$$\sum_{a_{-t}} \sigma(a)(v_t(a_{-t}, b_t) - v_t(a)) \;\leq\; \sum_{(a_{-t},s)} \sigma(a) \sum_{i \in t} \zeta_i(a, s)(\Pr(s|a_{-t}, b_t, \rho_t) - \Pr(s|a)).$$

Strong enforcement requires that no subset of agents can profitably deviate after coordinating their information even if they can commit to sharing their information non-strategically. This makes strong enforceability especially "strong."

We now derive the detection requirement implied by strong enforceability. Given a nonempty subset of agents $t \subset I$, a *multilateral deviation plan* for $t$ is any measure-valued map $\alpha_t : A_t \to \Delta(A_t \times R_t)$, where $A_t \times R_t = \prod_{i \in t} A_i \times R_i$. Intuitively, a multilateral deviation plan $\alpha_t$ has the agents in $t$ coordinate their deviations contingent on all recommendations to members of $t$. A multilateral deviation plan $\alpha_t$ is called *disobedient* if $\alpha_t(b_t, \rho_t|a_t) > 0$ for some $(a_t, b_t, \rho_t)$ such that $a_t \neq b_t$. It is called *detectable* if $\Pr(\sigma) \neq \Pr(\sigma, \alpha_t)$ for some $\sigma \in \Delta(A)$.

A *coalitional deviation plan* by agent $i$ is a profile of multilateral deviation plans $\alpha^i = \{\alpha_t : t \ni i\}$, one for each coalition to which $i$ may belong. It is called *disobedient* if $\alpha_t$ is disobedient for some coalition $t \ni i$. It is called *detectable* if $\Pr(\sigma) \neq \Pr(\sigma, \alpha^i)$ for some $\sigma \in \Delta(A)$, where $\Pr(\sigma, \alpha^i) := \sum_{t \ni i} \sum_{(a,b_t,\rho_t)} \sigma(a) \Pr(a_{-t}, b_t, \rho_t) \alpha_t(b_t, \rho_t|a_t)$. Intuitively, a coalitional deviation plan for an agent $i$ is a profile of multilateral deviation plans involving $i$. It is undetectable if regardless of the correlated strategy $\sigma$, even if some multilateral deviation plan $\alpha_t$ is detectable, there is another multilateral deviation plan $\alpha_{t'}$ with $i \in t \cap t'$ that "undoes" the change in probability from $\alpha_t$. Therefore, even if every disobedient multilateral deviation plan is detectable, it is possible that some disobedient coalitional deviation plan remains undetectable.

**Definition 9 (DCD).** A monitoring technology Pr *detects coalitional disobedience* (DCD) if every disobedient coalitional deviation plan is detectable.

The next result characterizes strong enforcement as detection of coalitional deviations. It is argued similarly to previous ones, so its proof is omitted.

**Theorem 13.** *A monitoring technology detects coalitional disobedience if and only if any team with any profile of coalitional utility functions can approximately strongly enforce every correlated strategy with secret contracts.*

# 5 Conclusion

Secret contracts emphasize that—as part of a team's economic organization—it may be beneficial for private information to be allocated differently across individuals in order for the right incentives to be provided. This remains true even if the team starts without informational asymmetry. Secret contracts effectively subject contractual deviations to "tailored monitoring," making monitors de facto auditors. Heuristically, secret contracts allow for monitoring to *follow* deviations.

We have provided above arguably weak notions of detection to characterize contractual enforcement. With budget-balanced contracts, we have shown that the appropriate notion of detection is attribution, which may be crudely interpreted as saying that "guilty until proven innocent" is—at least informationally—less costly than "innocent until proven guilty" as a principle for incentive provision.

# A Proofs

*Corollary 1.* Fix any $\sigma \in \Delta(A)$. By the Alternative Theorem, a monitoring technology Pr satisfies DUD-$\sigma$ if and only if there is a signal-contingent scheme $\zeta : I \times S \to \mathbb{R}$ such that

$$\forall i \in I, a_i \in B_i, (b_i, \rho_i) \in A_i \times R_i, \qquad 0 \leq \sum_{(a_{-i}, s)} \sigma(a)\zeta_i(s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

with a strict inequality if $a_i \neq b_i$, where $B_i = \{a_i \in A_i : \exists a_{-i} \text{ s.t. } \sigma(a) > 0\}$. Call this dual condition PSI-$\sigma$. By scaling $\zeta$ appropriately, PSI-$\sigma$ clearly implies that any deviation gains can be outweighed by contractual losses. Conversely, if DUD-$\sigma$ fails then $\Pr(\sigma) = \Pr(\sigma, \alpha_i)$ for some deviation plan $\alpha_i$ with $\alpha_i(b_i, \rho_i|a_i) > 0$ for some $a_i \in B_i$, and $b_i \neq a_i$. For all $a_{-i}$, let $0 = v_i(a) < v_i(a_{-i}, b_i) = 1$. Now $\sigma$ cannot be enforced by any $\zeta : I \times S \to \mathbb{R}$, since $\sum_{(b_i,\rho_i)} \alpha_i(b_i, \rho_i|a_i) \sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) > \sum_s \zeta_i(s)(\Pr(s|\sigma, \alpha_i) - \Pr(s|\sigma)) = 0$, being a convex combination of incentive constraints, must violate at least one. □

*Theorem 2.* By Lemma B.1, DUD is implied by *conic independence*

$$\forall (i, a_i, s_i), \qquad \Pr(a_i, s_i) \notin \text{cone}\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\}.$$

This is in turn implied by *linear* independence, or full row rank, for all $i$, of the matrix with $|A_i|\,|S_i|$ rows, $|A_{-i}|\,|S_{-i}|$ columns and entries $\Pr(a_i, s_i)(a_{-i}, s_{-i}) = \Pr(s|a)$. Since the set of full rank matrices is generic, this full row rank is generic if $|A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}|$ if $|S_i| > 1$

and $|S_{-i}| > 1$. If $|S_i| = 1$, adding with respect to $s_{-i}$ for each $a_{-i}$ yields column vectors equal to $(1, \ldots, 1) \in \mathbb{R}^{A_i}$. This leaves $|A_{-i}| - 1$ linearly dependent columns. Eliminating them, genericity requires that for every $i$,

$$|A_i| = |A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}| - (|A_{-i}| - 1) = |A_{-i}| \times (|S_{-i}| - 1) + 1.$$

Similarly, there are $|A_i| - 1$ redundant row vectors when $|S_{-i}| = 1$. Since the intersection of finitely many generic sets is generic, DUD is generic if all these conditions hold. □

*Corollary 3.* Consider the following primal problem: Find a feasible $\xi$ to solve

$$\forall (i, a_i, b_i),\ 0 \leq \sum_{(a_{-i}, s)} \xi_i(a, s)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)),\ \text{and}\ \forall (a, s),\ \sum_{i \in I} \xi_i(a, s) = K(a, s).$$

The dual of this problem is given by

$$\inf_{\lambda \geq 0, \eta} \sum_{(a,s)} \eta(a, s)K(a, s)\ \ \text{s.t.}\ \forall (i, a, s),\ \sum_{b_i \in A_i} \lambda_i(a_i, b_i)(\Pr(s|a_{-i}, b_i) - \Pr(s|a))\ =\ \eta(a, s).$$

If CEB is satisfied, then the value of the primal equals 0 for any $K : A \times S \to \mathbb{R}$. By the Strong Duality Theorem, the value of the dual is also 0 for any $K : A \times S \to \mathbb{R}$. Therefore, any $\eta$ satisfying the constraint for some $\lambda$ must be 0 for all $(a, s)$, so NOC is satisfied.

For necessity, if NOC is satisfied then the value of the dual is always 0 for any $K : A \times S \to \mathbb{R}$. By strong duality, the value of the primal is also 0 for any $K$. Therefore, given $K$, there is a feasible primal solution $\xi_i(a, s)$ that satisfies all primal constraints, and CEB holds. □

*Theorem 5.* Given the ordering of agents in the main text, if $k > 0$ permute agent $k$ with agent 1 and consider the following block matrix (blank spaces denote blocks of zeros).

$$Q\ =\ \begin{bmatrix} Q_1 & Q_1 & Q_1 & Q_1 & Q_1 \\ -Q_2 & & & & \\ & -Q_3 & & & \\ & & \cdots & -Q_{n-1} & \\ & & & & -Q_n \end{bmatrix}$$

where $Q_i$ is the matrix with $(|A_i|\,|S_i|)^2$ rows and $|A|\,|S|$ columns defined pointwise by

$$Q_i(a_i, s_i, b_i, t_i)(\widehat{a}, \widehat{s}) = \begin{cases} \Pr(\widehat{s}_{-i}, t_i|\widehat{a}_{-i}, b_i) & \text{if } (a_i, s_i) = (\widehat{a}_i, \widehat{s}_i) \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma B.2, IOA is satisfied if

$$\lambda Q = 0\ \ \text{and}\ \ \lambda \geq 0\ \ \Rightarrow\ \ \lambda_i(a_i, s_i, b_i, t_i) = 0\ \ \text{whenever}\ \ a_i \neq b_i,$$

which in turn is implied by $Q$ having full row rank.

Note that one row of $Q$ is redundant because for every agent $i > 1$,

$$\forall(\widehat{a},\widehat{s}), \qquad \sum_{(a_1,s_1)} Q_1(a_1,s_1,a_1,s_1)(\widehat{a},\widehat{s}) \;=\; \sum_{(a_i,s_i)} Q_i(a_i,s_i,a_i,s_i)(\widehat{a},\widehat{s}).$$

There may also be redundant column vectors. If $k > 1$, fix any agent $i \leq k$ with $i > 1$ and any $(a_1, a_i) \in A_1 \times A_i$. Then, for any $\widehat{a}$ such that $\widehat{a}_1 = a_1$ and $\widehat{a}_i = a_i$,

$$\forall(b_1,b_i), \qquad \sum_{\widehat{s}} Q_1(a_1,b_1)(\widehat{a},\widehat{s}) \;=\; 1 \qquad \text{and} \qquad \sum_{\widehat{s}} Q_i(a_i,b_i)(\widehat{a},\widehat{s}) \;=\; 1.$$

Therefore there are $|A_{-1,i}| - 1$ redundant columns for each $(a_1, a_i)$. If $|S_{-n}| = 1$ a similar argument shows that there are $(|A_n|\,|S_n|)^2\,|A_n| - 1$ additional redundant rows.

Therefore, $Q$ is generically full row rank if (1) the conditions for Theorem 2 are satisfied so that generically every $Q_i$ has full row rank and (2) the number of (non-redundant) rows is less than the number of (non-redundant) columns, i.e.,

$$\begin{aligned}\sum_{i=1}^{n}(|A_i|\,|S_i|)^2 - 1 - \chi_n(|A_n|^2\,|S_n| - 1) \;&\leq\; (n-1)\,|A|\,|S| - |A_1|\sum_{i=2}^{k-1}|A_i|\,(|A_{-1,i}| - 1)\\ &\leq\; (n-1)\,|A|\,|S| - (k-1)\,|A| + |A_1|\sum_{i=2}^{k-1}|A_i|\,,\end{aligned}$$

where $\chi_n = 1$ if $|S_{-n}| = 1$ and 0 otherwise. This inequality completes the proof. □

*Corollary 4.* By the Alternative Theorem, Pr satisfies $\text{DUD}_B^B$ if and only if it satisfies $\text{PSI}_B^B$, i.e., there exists a scheme $\xi : I \times A \times S \to \mathbb{R}$ such that $\xi_i(a,s) = 0$ if $a \notin B$ and

$$\forall i \in I, a_i \in B_i, b_i \in A_i, \rho_i \in R_i, \qquad 0 \;\leq\; \sum_{(a_{-i},s)} \xi_i(a,s)(\Pr(s|a_{-i},b_i,\rho_i) - \Pr(s|a)),$$

with a strict inequality whenever $a_i \neq b_i$. Replacing $\xi_i(a,s) = \sigma(a)\zeta_i(a,s)$ for some (or equivalently any) correlated strategy $\sigma$ with $\text{supp}\,\sigma = B$, this is equivalent to there being, for every profile of utility functions, an appropriate rescaling of $\zeta$ that satisfies (∗). □

*Theorem 6.* For necessity, if $\text{DUD}_B$ fails then there is a $B$-disobedient, undetectable deviation plan $\alpha_i$. Therefore, $\alpha_i(b_i,\rho_i|a_i) > 0$ for some $a_i \in B_i$, $b_i \neq a_i$ and $\rho_i \in R_i$. Letting $v_i(a_{-i},b_i) < v_i(a)$ for every $a_{-i}$, clearly no correlated strategy with positive probability on $a_i$ is approximately enforceable. Sufficiency follows by Lemmata B.3, B.4 and B.10. □

*Theorem 7.* Consider the family of linear programs below indexed by $z \in [0,\infty)^I$.

$$\max_{\varepsilon \geq 0, \xi} \; -\sum_{(i,a_i)} \varepsilon_i(a_i) \qquad \text{s.t.} \qquad \forall (i,a,s), \quad -\sigma(a) z_i \;\leq\; \xi_i(a,s) \;\leq\; \sigma(a) z_i,$$

$$\forall (i,a_i,b_i,\rho_i), \qquad \sum_{a_{-i}} \sigma(a) \Delta v_i(a,b_i) - \sum_{a_{-i}} \xi_i(a) \cdot \Delta \Pr(a,b_i,\rho_i) \;\leq\; \varepsilon_i(a_i),$$

where $\Delta v_i(a,b_i) := v_i(a_{-i},b_i) - v_i(a)$ and $\Delta \Pr(a,b_i,\rho_i) := \Pr(a_{-i},b_i,\rho_i) - \Pr(a)$. Given $z \geq 0$, the primal problem above looks for a scheme $\xi$ adapted to $\sigma$ (i.e., such that $\xi_i(a,s) = 0$ whenever $\sigma(a) = 0$) that minimizes the burden $\varepsilon_i(a_i)$ of relaxing incentive constraints. By construction, $\sigma$ is enforceable with transfers bounded by $z$ if and only if there is a feasible $\xi$ with $\varepsilon_i(a_i) = 0$ for all $(i,a_i)$, i.e., the value of the problem is zero. Since $\sigma$ is assumed enforceable, such $z$ exists. The dual of this problem is:

$$\min_{\alpha,\beta \geq 0} \sum_{(i,a)} \sigma(a)[z_i \sum_{s \in S} \sigma(a)(\beta_i^+(a,s) + \beta_i^-(a,s)) - \Delta v_i(a,\alpha_i)] \qquad \text{s.t.}$$

$$\forall (i,a_i), \qquad \sum_{(b_i,\rho_i)} \alpha_i(b_i,\rho_i|a_i) \;\leq\; 1,$$

$$\forall i \in I, a \in \text{supp } \sigma, s \in S, \qquad \Delta \Pr(s|a,\alpha_i) \;=\; \beta_i^+(a,s) - \beta_i^-(a,s).$$

Since $\beta_i^{\pm}(a,s) \geq 0$, it follows easily that $\beta_i^+(a,s) = \max\{\Delta \Pr(s|a,\alpha_i), 0\}$ and $\beta_i^-(a,s) = \min\{\Delta \Pr(s|a,\alpha_i), 0\}$. Hence, $\beta_i^+(a,s) + \beta_i^+(a,s) = |\Delta \Pr(s|a,\alpha_i)|$. Since $\|\Delta \Pr(a,\alpha_i)\| = \sum_s |\Delta \Pr(s|a,\alpha_i)|$, the dual is now equivalent to

$$V_\sigma(z) \;=\; \max_{\alpha \geq 0} \sum_{(i,a)} \sigma(a)(\Delta v_i(a,\alpha_i) - z \,\|\Delta \Pr(a,\alpha_i)\|) \;\text{ s.t. }\; \forall (i,a_i), \sum_{(b_i,\rho_i)} \alpha_i(b_i,\rho_i|a_i) \leq 1.$$

Adding mass to $\alpha_i(a_i,\tau_i|a_i)$ if necessary, without loss $\alpha_i$ is a deviation plan, proving $(i)$.

To prove $(ii)$, the first sentence is obvious. The second follows by Corollary 5: if $\sigma$ is not enforceable then a $\sigma$-profitable, supp $\sigma$-undetectable plan $\alpha_i$ exists, so $V_\sigma(z) > 0$ for all $z$.

For $(iii)$, if $\sigma$ is not enforceable then there is a $\sigma$-profitable, supp $\sigma$-undetectable deviation plan $\alpha_i^*$. Approaching $\alpha_i^*$ from $\mathscr{F}_i$ (e.g., with mixtures of $\alpha_i^*$ and a fixed plan in $\mathscr{F}_i$), the denominator defining $\overline{z}_i$ tends to zero whilst the numerator tends to a positive amount, so $\overline{z}_i$ is unbounded. Conversely, suppose $\sigma$ is enforceable. If the sup defining $\overline{z}_i$ is attained, we are done. If not, it is approximated by a sequence of supp $\sigma$-detectable deviation plans that converge to a supp $\sigma$-undetectable one. Since $\sigma$ is enforceable, the limit is unprofitable. Let

$$F_i^\sigma(\delta) \;:=\; \min_{\lambda_i \geq 0} \sum_{a \in A} \sigma(a) \,\|\Delta \Pr(a,\lambda_i)\| \;\text{ s.t. }\; \Delta v_i(\sigma,\lambda_i) \;\geq\; \delta.$$

Since every $\sigma$-profitable deviation plan is detectable by Corollary 5, it follows that $F_i^\sigma(\delta) > 0$ for all $\delta > 0$, and $\overline{z}_i = (\lim_{\delta \downarrow 0} F_i^\sigma(\delta)/\delta)^{-1}$. Hence, it suffices to show $\lim_{\delta \downarrow 0} F_i^\sigma(\delta)/\delta > 0$.

To this end, by adding variables like $\beta$ above, the dual problem for $F_i^\sigma$ is equivalent to:

$$F_i^\sigma(\delta) \;=\; \max_{\varepsilon\geq 0, x_i} \; \varepsilon\delta \;\text{ s.t. }\; \forall(a,s), \quad -1 \;\leq\; x_i(a,s) \;\leq\; 1,$$
$$\forall(a_i,b_i,\rho_i), \qquad \sum_{a_{-i}} \sigma(a)(\varepsilon\Delta v_i(a,b_i) - x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i)) \;\leq\; 0.$$

Since $\sigma$ is enforceable, there is a feasible solution to this dual $(\varepsilon, x_i)$ with $\varepsilon > 0$. Hence, $F_i^\sigma(\delta) \geq \varepsilon\delta$ for all $\delta > 0$, therefore $\lim_{\delta\downarrow 0} F_i^\sigma(\delta)/\delta > 0$, as claimed.

To prove $(iv)$, suppose that $\overline{z}_i < \infty$ for all $i$. We claim $V_\sigma(\overline{z}) = 0$. Indeed, given $\alpha_i^* \in \mathscr{F}_i$ for all $i$, substituting the definition of $\overline{z}_i$ into the objective of the minimization in $(i)$,

$$\sum_{i\in I}\Delta v_i(\sigma,\alpha_i^*) - \sum_{(i,a)}\sigma(a)\sup_{\alpha_i\in\mathscr{F}_i}\{\frac{\max\{\Delta v_i(\sigma,\alpha_i),0\}}{\sum_a \sigma(a)\,\|\Delta\Pr(a,\alpha_i)\|}\}\,\|\Delta\Pr(a,\alpha_i^*)\| \;\leq\; 0.$$

If $\alpha_i^* \notin \mathscr{F}_i$ then, since $\sigma$ is enforceable, every supp $\sigma$-undetectable deviation plan is unprofitable, so again the objective is non-positive, hence $V_\sigma(\overline{z}) = 0$. Clearly, $V_\sigma$ decreases with $z$, so it remains to show that $V_\sigma(\overline{z}) > 0$ if $z_i < \overline{z}_i$ for some $i$. But by definition of $\overline{z}$, there is a deviation plan $\alpha_i^*$ with $\Delta v_i(\sigma,\alpha_i^*)/\sum_a \sigma(a)\,\|\Delta\Pr(a,\alpha_i^*)\| > z_i$, so $V_\sigma(z) > 0$. □

*Theorem 8.* For sufficiency, suppose that $\sigma$ is approximately enforceable, so there is a sequence $\{\sigma^m\}$ such that $\sigma^m$ is enforceable for every $m$ and $\sigma^m \to \sigma$. Without loss, assume that supp $\sigma^m \supset$ supp $\sigma$ for all $m$. If $\sigma^m = \sigma$ for all large $m$ then $\sigma$ is enforceable and the condition of Theorem 8 is fulfilled with $\mu = \sigma$, so suppose not. If there exists $m$ and $m'$ such that $\sigma^m = p\sigma^{m'} + (1-p)\sigma$ then incentive compatibility with respect to $m$ yields that $\sum_{a_{-i}}\sigma^m(a)\Delta v_i(a,\alpha_i) \leq \sum_{a_{-i}}\sigma^m(a)\zeta_i^m(a)\cdot\Delta\Pr(a,\alpha_i) \leq \sum_{a_{-i}}\sigma^m(a)\overline{z}\,\|\Delta\Pr(a,\alpha_i)\|$ for every $\alpha_i$, where $\overline{z} = \max_{(i,a,s)}|\zeta_i^m(a,s)|$. For large $m'$, $\sigma^{m'}$ is sufficiently close to $\sigma$ that if $\alpha_i$ is $\sigma$-profitable then $\sum_{a_{-i}}\sigma^{m'}(a)\Delta v_i(a,\alpha_i) > 0$, so $\alpha_i$ is detectable.

If there does not exist $m$ and $m_1$ such that $\sigma^m = p\sigma^{m_1} + (1-p)\sigma$ then there exists $\sigma^{m_2}$ such that its distance from $\sigma$ is less than the positive minimum distance between $\sigma$ and the affine hull of $\{\sigma^m, \sigma^{m_1}\}$. Therefore, the lines generated by $\sigma^m$ and $\sigma^{m_1}$ and $\sigma^{m_1}$ and $\sigma^{m_2}$ are not collinear. Proceeding inductively, pick $C = \{\sigma^{m_1}, \ldots, \sigma^{m_{|A|}}\}$ such that its affine space is full-dimensional in $\Delta(A)$. Since we are assuming that $\sigma$ is not enforceable, it lies outside conv $C$. Let $\widehat{\sigma} = \sum_k \sigma^{m_k}/|A|$ and $B_\varepsilon(\widehat{\sigma})$ be the open $\varepsilon$-ball around $\widehat{\sigma}$ for some $\varepsilon > 0$. By construction, $B_\varepsilon(\widehat{\sigma}) \subset$ conv $C$ for $\varepsilon > 0$ sufficiently small, so there exists $\widehat{\sigma}' \in B_\varepsilon(\widehat{\sigma})$ such that $p\widehat{\sigma} + (1-p)\sigma = \widehat{\sigma}'$ for some $p$ such that $0 < p < 1$. Now we can apply the argument from the previous paragraph, so the condition of Theorem 8 holds.

For necessity, if $\sigma$ is not approximately enforceable then $1 \geq V_\sigma(z) \geq C > 0$ for every $z$,

where $V_\sigma$ is defined in Lemma B.3. Let $(\lambda^z, \mu^z)$ solve $V_\sigma(z)$ for every $z$. Given $\mu \in \Delta(A)$,

$$C \;\le\; V_\sigma(z) \;\le\; 1 + \sum_{(i,a)} \Delta v_i(\mu, \lambda_i^z) - z \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\|.$$

If the condition of Theorem 8 holds then $\sum_{(i,a)} \Delta v_i(\mu, \lambda_i^z) < \overline{z} \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\|$ and $\sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\| > 0$, since there must exist $i$ such that $\lambda_i^\sigma$ is $\sigma$-profitable. Hence, $C \le 1 + (\overline{z} - z) \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\|$, i.e., $z - \overline{z} \le (1-c) / \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\|$. This inequality must hold for every $z$, therefore $\sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^z) \right\| \to 0$ as $z \to \infty$. But this contradicts Lemma B.11, since $\sum_i \Delta v_i(\sigma, \lambda_i) \ge C$, completing the proof. □

*Theorem 9.* Enforcing an arbitrary correlated strategy $\sigma$ subject to budget balance and participation reduces to finding transfers $\zeta$ to solve the following family of linear inequalities:

$$\begin{aligned}
\forall (i, a_i, b_i, \rho_i), \qquad \Delta v_i(\sigma, b_i) \;&\le\; \sum_{a_{-i}} \sigma(a) \zeta_i(a) \cdot \Delta \Pr(a, b_i, \rho_i), \\
\forall (a, s), \qquad \sum_{i=1}^{n} \zeta_i(a, s) \;&=\; 0, \\
\forall i \in I, \qquad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a, s) \Pr(s|a) \;&\ge\; 0.
\end{aligned}$$

The dual of this metering problem subject to participation is:

$$\max_{\lambda, \pi \ge 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \pi_i v_i(\sigma) \;\; \text{s.t. } \forall (i, a, s), \quad \sigma(a) \Delta \Pr(s|a, \lambda_i) \;=\; \eta(a, s) + \pi_i \sigma \Pr(s|a)$$

where $\pi_i$ is a multiplier for agent $i$'s participation constraint and $v_i(\sigma) = \sum_a \sigma(a) v_i(a)$. Adding the dual constraints with respect to $s \in S$, it follows that $\pi_i = \pi$ does not depend on $i$. Redefining $\eta(a, s)$ as $\eta(a, s) + \pi \Pr(s|a)$, the set of feasible $\lambda \ge 0$ is the same as without participation constraints. Since $\sum_i v_i(a) \ge 0$ for all $a$, the dual is maximized by $\pi = 0$. □

*Theorem 11.* We just prove the result with budget balance (without budget balance it follows similarly). Let $z = (z_1, \ldots, z_n)$ be a vector of liability limits for each agent. Enforcing $\sigma$ subject to budget balance and one-sided limited liability reduces to finding $\zeta$ such that

$$\begin{aligned}
\forall (i, a_i, b_i, \rho_i), \qquad \Delta v_i(\sigma, b_i) \;&\le\; \sum_{a_{-i}} \sigma(a) \zeta_i(a) \cdot \Delta \Pr(a, b_i, \rho_i), \\
\forall (a, s), \qquad \sum_{i=1}^{n} \zeta_i(a, s) \;&=\; 0, \\
\forall (i, a, s), \qquad \zeta_i(a, s) \;&\le\; z_i.
\end{aligned}$$

The dual of this metering problem subject to one-sided limited liability is given by:

$$\max_{\lambda, \beta \ge 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \sum_{(i,a,s)} \beta_i(a, s) z_i \;\; \text{s.t. } \forall (i, a, s), \;\; \sigma(a) \Delta \Pr(s|a, \lambda_i) = \eta(a, s) + \beta_i(a, s),$$

where $\beta_i(a,s)$ is a multiplier on the liability constraint for agent $i$ at $(a,s)$. Adding the dual equations with respect to $s$ implies $-\sum_s \beta_i(a,s) = \sum_s \eta(a,s)$ for all $(i,a)$. Therefore,

$$-\sum_{(i,s)} \beta_i(a,s) z_i \;=\; \sum_{(i,s)} \eta(a,s) z_i \;=\; \widehat{z} \sum_{s \in S} \eta(a,s),$$

where $\widehat{z} = \sum_i z_i$, so we may eliminate $\beta_i(a,s)$ from the dual and get the equivalent problem:

$$\max_{\lambda \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) + \widehat{z} \sum_{(a,s)} \eta(a,s) \;\text{ s.t. }\; \forall (i,a,s),\; \sigma(a) \Delta \Pr(s|a,\lambda_i) \;\geq\; \eta(a,s).$$

Any two liability profiles $z$ and $z'$ with $\widehat{z} = \widehat{z}'$ lead to this dual with the same value. □

*Theorem 12.* Enforcing a correlated strategy $\sigma$ subject to budget balance, participation and one-sided limited liability reduces to finding a solution $\zeta$ to the following linear system:

$$\begin{aligned}
\forall (i, a_i, b_i, \rho_i), \qquad \Delta v_i(\sigma, b_i) \;&\leq\; \sum_{a_{-i}} \sigma(a) \zeta_i(a) \cdot \Delta \Pr(a, b_i, \rho_i), \\
\forall (a,s), \qquad \sum_{i \in I} \zeta_i(a,s) \;&=\; 0, \\
\forall i \in I, \quad \sum_{a \in A} \sigma(a)(v_i(a) - \zeta_i(a) \cdot \Pr(a)) \;&\geq\; 0, \\
\forall (i,a,s), \qquad \zeta_i(a,s) \;&\leq\; z_i.
\end{aligned}$$

The first family of constraints imposes incentive compatibility, the second budget balance, the third individual rationality, and the last corresponds to one-sided limited liability. The dual of this metering problem is given by the following program, where $\lambda$, $\eta$, $\pi$ and $\beta$ represent the respective multipliers on each of the primal constraints.

$$\begin{aligned}
\max_{\lambda, \pi, \beta \geq 0, \eta} \; \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \sum_{i \in I} \pi_i v_i(\sigma) - \sum_{(i,a,s)} \beta_i(a,s) z_i \quad \text{s.t.} \\
\forall (i,a,s), \qquad \sigma(a) \Delta \Pr(s|a,\lambda_i) \;=\; \eta(a,s) + \pi_i \sigma(a) \Pr(s|a) + \beta_i(a,s).
\end{aligned}$$

Adding the dual constraints with respect to $s \in S$, it follows that

$$-\sum_{(a,s)} \beta_i(a,s) = \sum_{(a,s)} \eta(a,s) + \pi_i \;=\; \widehat{\eta} + \pi_i$$

where $\widehat{\eta} := \sum_{(a,s)} \eta(a,s)$. After substituting and eliminating $\beta$, the dual is equivalent to

$$\begin{aligned}
V \;:=\; \max_{\lambda, \pi \geq 0, \eta} \; \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \widehat{z} \quad \text{s.t.} \\
\forall (i,a,s), \qquad \sigma(a) \Delta \Pr(s|a,\lambda_i) \;\geq\; \eta(a,s) + \pi_i \sigma(a) \Pr(s|a).
\end{aligned}$$

It is clear that Theorems 9 and 11 follow almost immediately from this linear program. Now, $\sigma$ is enforceable if and only if $V = 0$, i.e., if and only if for any dual-feasible $(\lambda, \pi, \eta)$ such that $\sum_i \Delta v_i(\sigma, \lambda_i) > 0$, we have that

$$\sum_{i \in I} \Delta v_i(\sigma, \lambda_i) \;\leq\; \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta}\widehat{z}.$$

Finally, since the dual objective is decreasing in $\eta$, an optimal solution for $\eta$ must solve

$$\widehat{\eta} \;=\; \max_{s \in S} \{\pi_i \Pr(s|a) - \Pr(s|a, b_i, \rho_i)\}.$$

This completes the proof. □

*Corollary 7.* Given the dual problem from the proof of Theorem 12, the first statement follows because if $v_i(\sigma) \geq z_i$ then the objective function is decreasing in $\pi_i$ and reducing $\pi_i$ relaxes the dual constraints. The second statement follows by rewriting the objective as

$$\sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \sum_{i \in I \setminus t} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \sum_{i \in I} z_i,$$

where $t$ is the set of agents whose participation constraint won't bind ($\pi_i^* = 0$ for $i \in t$). □

# B Lemmata

**Lemma B.1.** *A monitoring technology satisfies DUD if*

$$\forall (i, a_i, s_i), \qquad \Pr(a_i, s_i) \notin \text{cone}\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\},$$

*where* cone *stands for the set of positive linear combinations of* $\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\}$.

*Proof.* If DUD fails then there exists $\alpha_i$ such that $\alpha_i(b_i, \rho_i | a_i) > 0$ for some $a_i \neq b_i$ and

$$\begin{aligned} \forall (a, s), \qquad \Pr(s|a) \;&=\; \sum_{(b_i, \rho_i)} \sum_{t_i \in \rho_i^{-1}(s_i)} \alpha_i(b_i, \rho_i | a_i) \Pr(s_{-i}, t_i | a_{-i}, b_i) \\ &=\; \sum_{(b_i, t_i)} \sum_{\{\rho_i : \rho_i(t_i) = s_i\}} \alpha_i(b_i, \rho_i | a_i) \Pr(s_{-i}, t_i | a_{-i}, b_i). \end{aligned}$$

Write $\lambda_i(a_i, s_i, b_i, t_i) := \sum_{\{\rho_i : \rho_i(t_i) = s_i\}} \alpha_i(b_i, \rho_i | a_i)$. By construction, $\lambda_i(a_i, s_i, b_i, t_i) \geq 0$ is strictly positive for some $a_i \neq b_i$ and satisfies

$$\forall (i, a, s), \qquad \Pr(s|a) \;=\; \sum_{(b_i, t_i)} \lambda_i(a_i, s_i, b_i, t_i) \Pr(s_{-i}, t_i | a_{-i}, b_i).$$

Without loss, $\lambda_i(a_i, s_i, a_i, s_i) = 0$ for some $(a_i, s_i)$. Indeed, if $\lambda_i(a_i, s_i, a_i, s_i) = 1$ for all $(a_i, s_i)$, then the equation above is violated because $\alpha_i$ is disobedient by hypothesis and probabilities are non-negative. If $\lambda_i(a_i, s_i, a_i, s_i) \neq 1$ then subtract $\lambda_i(a_i, s_i, a_i, s_i) \Pr(s|a)$ from both sides of the equation and divide by $1 - \lambda_i(a_i, s_i, a_i, s_i)$. Therefore, $\Pr(a_i, s_i) \in \text{cone}\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\}$ for some $(a_i, s_i)$. □

**Lemma B.2.** *A monitoring technology satisfies IOA if*

$$\forall (i, j, a, s), \quad \sum_{(b_i, t_i)} \lambda_i(a_i, s_i, b_i, t_i) \Pr(s_{-i}, t_i | a_{-i}, b_i) = \sum_{(b_j, t_j)} \lambda_j(a_j, s_j, b_j, t_j) \Pr(s_{-j}, t_j | a_{-j}, b_j)$$

*and $\lambda \geq 0$ implies that for every agent $i$, $\lambda_i(a_i, s_i, b_i, t_i) = 0$ whenever $a_i \neq b_i$.*

*Proof.* If IOA fails then there exists $\lambda \geq 0$ and $\eta$ such that $\lambda_i(a_i, b_i, \rho_i) > 0$ for some $a_i \neq b_i, i$ and

$$\forall (i, a, s), \qquad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) \;=\; \eta(a, s),$$

By adjusting $\lambda_i(a_i, a_i, \tau_i)$ if necessary, assume without loss that $\sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i) = \Lambda$ for some constant $\Lambda$, for every $i$ and $a_i$. Therefore,

$$\forall (i, a, s), \qquad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i) \Pr(s|a_{-i}, b_i, \rho_i) \;=\; \Lambda \Pr(s|a) - \eta(a, s).$$

The result follows now by the same argument as for Lemma B.1 above. □

**Lemma B.3.** *Consider the following linear program.*

$$\begin{aligned}
V_\sigma(z) \;&:=\; \min_{\mu \geq 0, p, \xi} \; p \;\text{ s.t. }\; \sum_{a \in A} \mu(a) \;=\; p, \\
\forall (i, a, s), \qquad & -(\mu(a) + (1-p)\sigma(a))z \;\leq\; \xi_i(a, s) \;\leq\; (\mu(a) + (1-p)\sigma(a))z, \\
\forall (i, a_i, b_i, \rho_i), \qquad & \sum_{a_{-i}} (\mu(a) + (1-p)\sigma(a)) \Delta v_i(a, b_i) \;\leq\; \sum_{a_{-i}} \xi_i(a) \cdot \Delta \Pr(a, b_i, \rho_i).
\end{aligned}$$

*The correlated strategy $\sigma$ is approximately enforceable if and only if $V_\sigma(z) \to 0$ as $z \to \infty$. The dual of the above linear program is given by the following problem:*

$$\begin{aligned}
V_\sigma(z) \;=\; \max_{\lambda \geq 0, \kappa} \; & \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - z \sum_{(i,a)} \sigma(a) \left\| \Delta \Pr(a, \lambda_i) \right\| \quad \text{s.t.} \\
\forall a \in A, \qquad & \kappa \;\leq\; \sum_{i \in I} \Delta v_i(a, \lambda_i) - z \sum_{i \in I} \left\| \Delta \Pr(a, \lambda_i) \right\|, \\
& \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - z \sum_{(i,a)} \sigma(a) \left\| \Delta \Pr(a, \lambda_i) \right\| \;=\; 1 + \kappa.
\end{aligned}$$

*Proof.* The first family of primal constraints require $\xi$ to be adapted to $\mu + (1-p)\sigma$, so for any $z$, $(\mu, p, \xi)$ solves the primal if and only if $\mu + (1-p)\sigma$ is exactly enforceable with $\xi$. (Since correlated equilibrium exists, the primal constraint set is clearly nonempty, and for finite $z$ it is also clearly bounded). The first statement now follows. The second statement follows by a lengthy but standard manipulation of the primal to obtain the above dual. □

**Lemma B.4.** *Consider the following family of linear programs indexed by $\varepsilon > 0$ and $z \geq 0$.*

$$\begin{aligned} F_\sigma^\varepsilon(z) \;:=\; \max_{\lambda \geq 0} \min_{\mu \in \Delta(A)} & \sum_{i \in I} \Delta v_i(\mu, \lambda_i) - z \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i) \right\| \text{ s.t.} \\ & \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - z \sum_{(i,a)} \sigma(a) \left\| \Delta \Pr(a, \lambda_i) \right\| \;\geq\; \varepsilon. \end{aligned}$$

*$F_\sigma^\varepsilon(z) \to -\infty$ as $z \to \infty$ for some $\varepsilon > 0$ if and only if $\sigma$ is approximately enforceable.*

*Proof.* The dual of the problem defining $F_\sigma^\varepsilon(z)$ is

$$\begin{aligned} F_\sigma^\varepsilon(z) \;=\; \min_{\delta, \mu \geq 0, x} \; -\delta\varepsilon \;\text{ s.t. }\; \sum_{a \in A} \mu(a) \;=\; 1, \\ \forall (i,a,s), \qquad -(\mu(a) + \delta\sigma(a))z \;\leq\; x_i(a,s) \;\leq\; (\mu(a) + \delta\sigma(a))z, \\ \forall (i, a_i, b_i, \rho_i), \qquad \sum_{a_{-i}} (\mu(a) + \delta\sigma(a)) \Delta v_i(a, b_i) \;\leq\; \sum_{a_{-i}} x_i(a) \cdot \Delta \Pr(a, b_i, \rho_i). \end{aligned}$$

Since clearly $\varepsilon > 0$ does not affect the dual feasible set, if $F_\sigma^\varepsilon(z) \to -\infty$ for some $\varepsilon > 0$ then there exists $z \geq 0$ such that $\delta > 0$ is feasible, and $\delta \to \infty$ as $z \to \infty$. Therefore, $F_\sigma^\varepsilon(z) \to -\infty$ for every $\varepsilon > 0$. If $V_\sigma(z) = 0$ for some $z$ we are done by monotonicity of $V_\sigma$. Otherwise, suppose that $V_\sigma(z) > 0$ for all $z > 0$. Let $(\lambda, \kappa)$ be an optimal dual solution for $V_\sigma(z)$ in Lemma B.3. By optimality, $\kappa = \min_{\mu \in \Delta(A)} \sum_i \Delta v_i(\mu, \lambda_i) - z \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i) \right\|$. Therefore, by the second dual constraint in $V_\sigma(z)$ of Lemma B.3,

$$V_\sigma(z) \;=\; 1 + \kappa \;=\; 1 + F_\sigma^{V_\sigma(z)}(z) \;=\; 1 - \delta V_\sigma(z),$$

where $\delta$ is an optimal solution to the dual with $\varepsilon = V_\sigma(z)$. Rearranging, $V_\sigma(z) = 1/(1+\delta)$. Finally, $F_\sigma^\varepsilon(z) \to -\infty$ as $z \to \infty$ if and only if $\delta \to \infty$, if and only if $V_\sigma(z) \to 0$. □

**Lemma B.5.** *Fix any $\varepsilon > 0$. If* Pr *satisfies $DUD_B$, where $B = \operatorname{supp} \sigma$, then for every $C \leq 0$ there exists $z \geq 0$ such that $G_\sigma(z) \leq C$, where $\Delta v_i(a_i)^* := \max_{(a_{-i}, b_i)} \{\Delta v_i(a, b_i)\}$,*

$$\begin{aligned} \Delta v_i(a_i, \lambda_i)^* \;:=\; \Delta v_i(a_i)^* \sum_{(a_i, b_i \neq a_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i), \;\text{ and} \\ G_\sigma(z) \;:=\; \max_{\lambda \geq 0} \sum_{(i,a)} \left\| \Delta v_i(a_i, \lambda_i) \right\| - z \sum_{(i,a)} \left\| \Delta \Pr(a, \lambda_i) \right\| \qquad \text{s.t.} \\ \forall i \in I, a_i \notin B_i, \; \lambda_i(a_i) \;=\; 0, \qquad \text{and} \qquad \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - z \sum_{(i,a)} \sigma(a) \left\| \Delta \Pr(a, \lambda_i) \right\| \;\geq\; \varepsilon. \end{aligned}$$

*Proof.* The dual of this problem is given by

$$\begin{aligned} G_\sigma(z) \;&=\; \min_{\delta\geq 0,x} \; -\delta\varepsilon \;\text{ s.t.} \\ \forall(i,a,s), \qquad -(1+\delta\sigma(a))z \;&\leq\; x_i(a,s) \;\leq\; (1+\delta\sigma(a))z, \\ \forall(i,a_i\in B_i,b_i,\rho_i), \qquad \sum_{a_{-i}} \delta\sigma(a)\Delta v_i(a,b_i) + \mathbf{1}_{\{a_i\neq b_i\}}\Delta v_i(a_i)^* \;&\leq\; \sum_{a_{-i}} x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i), \end{aligned}$$

where $\mathbf{1}_{\{b_i\neq a_i\}} = 1$ if $b_i \neq a_i$ and 0 otherwise. This problem looks almost exactly like the dual for $F^\varepsilon_\sigma(z)$ except that the incentive constraints are only indexed by $a_i \in B_i$. Now, $\mathrm{DUD}_B$ is equivalent to $\mathrm{PSI}_B$, i.e., there is an incentive scheme $x : I\times A\times S\to\mathbb{R}$ such that

$$\forall(i,a_i,b_i,\rho_i), \qquad 0 \;\leq\; \sum_{a_{-i}} x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i)$$

with a strict inequality whenever $a_i \in B_i$ and $a_i \neq b_i$. Hence, by scaling $x$ appropriately, there is a feasible dual solution with $\delta > 0$, so $G_\sigma(z) < 0$. Moreover, for any $\delta > 0$, clearly an $x$ exists with $\sum_{a_{-i}} \delta\sigma(a)\Delta v_i(a,b_i) + \mathbf{1}_{\{b_i\neq a_i\}}\Delta v_i(a_i)^* \leq \sum_{a_{-i}} x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i)$ on all $(i,a_i\in B_i,b_i,\rho_i)$ by $\mathrm{PSI}_B$, so there exists $z$ to make such $\delta$ feasible. In particular, $\delta \geq C/\varepsilon$ is feasible for some $z$, as required. □

**Lemma B.6.** *If* Pr *satisfies* $DUD_B$*, then there exists a finite* $z \geq 0$ *such that*

$$\forall i\in I, a_i\in B_i, \lambda_i\geq 0, \qquad \sum_{a_{-i}} \Delta v_i(a_i,\lambda_i)^* - z\,\|\Delta\Pr(a,\lambda_i)\| \;\leq\; 0.$$

*Proof.* Given $i$, $a_i \in B_i$, plug $\sigma(a) = 1/|A_{-i}|$ for all $a_{-i}$ in the proof of Theorem 7 (*iii*). □

Call $\lambda$ *extremely detectable* if $\lambda_i(a_i)$ cannot be written as a positive linear combination involving undetectable deviations (possibly mixed) for every $(i,a_i)$. Let $\mathscr{E}$ denote the set of all such extremely detectable $\lambda$.

**Lemma B.7.** *The set* $\mathscr{D}^e = \{\alpha\in\mathscr{E} : \forall(i,a_i), \sum_{(b_i,\rho_i)}\alpha_i(a_i,b_i,\rho_i) = 1\}$ *is compact.*

*Proof.* $\mathscr{D}^e$ is clearly a bounded subset of Euclidean space, so it remains to show that it is closed. Consider a sequence $\{\alpha^m\}\subset\mathscr{D}^e$ such that $\alpha^m\to\alpha^*$. For any $\alpha\in\mathscr{D}$, let

$$p^*(\alpha) \;:=\; \max_{0\leq p\leq 1,\alpha^i\in\mathscr{D}} \{p : \alpha^0 \text{ is undetectable, } p\alpha^0 + (1-p)\alpha^1 = \alpha\}.$$

This is a well-defined linear program with a compact constraint set and finite values, so $p^*$ is continuous in $\alpha$. By assumption, $p^*(\alpha^m) = 0$ for all $m$, so $p^*(\alpha^*) = 0$, hence $\alpha^*\in\mathscr{D}^e$. □

**Lemma B.8.** *Let* $\mathscr{D}^e$ *be the set of extremely detectable deviation plans.*

$$\gamma \;:=\; \min_{\alpha^e\in\mathscr{D}^e} \sum_{(i,a)} \|\Delta\Pr(a,\alpha^e_i)\| \;>\; 0.$$

*Proof.* If $\mathscr{D}^e = \emptyset$ then $\gamma = +\infty$. If not, $\mathscr{D}^e$ is compact by Lemma B.7, so there is no sequence $\{\alpha_i^{e,m}\} \subset \mathscr{D}^e$ with $\|\Delta \Pr(a, \alpha_i^{e,m})\| \to 0$ for all $(i,a)$ as $m \to \infty$, hence $\gamma > 0$. □

**Lemma B.9.** *Let $\mathscr{D}_i^e = \mathrm{proj}_i \mathscr{D}^e$. There exists a finite $z \geq 0$ such that*

$$\forall i \in I, a_i \notin B_i, \alpha_i^e \in \mathscr{D}_i^e, \qquad \sum_{a_{-i}} \Delta v_i(a_i, \alpha_i^e)^* - z \|\Delta \Pr(a, \alpha_i^e)\| \ \leq \ 0.$$

*Proof.* Let $\|\Delta v\| = \max_{(i,a,b_i)} |\Delta v_i(a, b_i)|$. If $z \geq \|\Delta v\| / \gamma$, with $\gamma$ as in Lemma B.8, then

$$\forall (i, a_i), \quad \sum_{a_{-i}} \Delta v_i(a_i, \alpha_i^e)^* - z \|\Delta \Pr(a, \alpha_i^e)\| \ \leq \|\Delta v\| - z \sum_{a_{-i}} \|\Delta \Pr(a, \alpha_i^e)\| \ \leq \ \|\Delta v\| - \frac{\|\Delta v\|}{\gamma} \gamma.$$

The right-hand side clearly equals zero, which establishes the claim. □

**Lemma B.10.** *Fix any $\varepsilon > 0$. If* Pr *satisfies $DUD_B$ then for every $C \leq 0$ there exists $z \geq 0$ such that for every $\lambda \geq 0$ with*

$$\sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - z \sum_{(i,a)} \sigma(a) \|\Delta \Pr(a, \lambda_i)\| \ \geq \ \varepsilon,$$

*there exists $\mu \in \Delta(A)$ such that*

$$W(\mu, \lambda) \ := \ \sum_{i \in I} \Delta v_i(\mu, \lambda_i) - z \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\| \ \leq \ C.$$

*Proof.* Rewrite $W(\mu, \lambda)$ by splitting it into three parts, $W_d(\mu, \lambda)$, $W_e(\mu, \lambda)$ and $W_u(\mu, \lambda)$:

$$\begin{aligned}
W_d(\mu, \lambda) \ &= \ \sum_{i \in I} \sum_{a_i \in B_i} \sum_{a_{-i}} \mu(a) (\Delta v_i(a, \lambda_i) - z \|\Delta \Pr(a, \lambda_i)\|) \\
W_e(\mu, \lambda) \ &= \ \sum_{i \in I} \sum_{a_i \notin B_i} \sum_{a_{-i}} \mu(a) (\Delta v_i(a, \lambda_i^e) - z \|\Delta \Pr(a, \lambda_i^e)\|), \\
W_u(\mu, \lambda) \ &= \ \sum_{i \in I} \sum_{a_i \notin B_i} \sum_{a_{-i}} \mu(a) (\Delta v_i(a, \lambda_i^u) - z \|\Delta \Pr(a, \lambda_i^u)\|),
\end{aligned}$$

and $\lambda = \lambda^e + \lambda^u$ with $\lambda^e$ extremely detectable, $\lambda^u$ undetectable. Since $\lambda^u$ is undetectable,

$$W_u(\mu, \lambda) \ = \ \sum_{i \in I} \sum_{a_i \notin B_i} \sum_{a_{-i}} \mu(a) \Delta v_i(a, \lambda_i^u)$$

Let $\mu^0(a) = 1/|A|$ for every $a$. By Lemma B.5, there exists $z$ with $W_d(\mu^0, \lambda) \leq C$ for every $\lambda$, and by Lemma B.9 there exists $z$ with $W_e(\mu^0, \lambda) \leq 0$ for every $\lambda$. Therefore, if $W_u(\mu^0, \lambda) \leq 0$ we are done. Otherwise, for every $i$ and $a_i, b_i \in A_i$, let $\mu_i^0(a_i) = 1/|A_i|$ and

$$\mu_i^1(b_i) \ := \ \sum_{(a_i, \rho_i)} \frac{\lambda_i^u(a_i, b_i, \rho_i)}{\sum_{(b_i', \rho_i')} \lambda_i^u(a_i, b_i', \rho_i')} \mu_i^0(a_i)$$

Iterate this rule to obtain a sequence $\{\mu_i^m\}$ with limit $\mu_i^\infty \in \Delta(A_i)$. By construction, $\mu_i^\infty$ is a $\lambda_i^u$-stationary distribution (Nau and McCardle, 1990; Myerson, 1997). Therefore, given any $a_{-i}$, the deviation gains for every agent equal zero, i.e.,

$$\sum_{(a_i,b_i,\rho_i)} \mu_i^\infty(a_i)\lambda_i^u(a_i,b_i,\rho_i)(v_i(a_{-i},b_i) - v_i(a)) \;=\; 0.$$

Let $\mu^m(a) := \prod_i \mu_i^m(a_i)$ for all $m$. By construction, $W_u(\mu^\infty,\lambda^u) = 0$. We will show that $W_d(\mu^\infty,\lambda) \le C$ and $W_e(\mu^\infty,\lambda) \le 0$. To see this, notice firstly that, since $\lambda_i^u$ is undetectable, for any other agent $j \neq i$, any $\lambda_j \ge 0$ and every action profile $a \in A$,

$$\|\Delta \Pr(a,\lambda_j)\| \;=\; \|\Delta \Pr(a,\lambda_i^u,\lambda_j)\| \;\le\; \|\Delta \Pr(a,\widehat{\lambda}_i^u,\lambda_j)\|,$$

where $\widehat{\lambda}_i^u(a_i,b_i,\tau_i) = \sum_{\rho_i} \lambda_i^u(a_i,b_i,\rho_i)$ and $\widehat{\lambda}_i^u(a_i,b_i,\rho_i) = 0$ for all $\rho_i \neq \tau_i$,

$$\Delta \Pr(a,\lambda_i^u,\lambda_j) = \sum_{(b_j,\rho_j)} \lambda_j(a_j,b_j,\rho_j) \sum_{(b_i,\rho_i)} \lambda_i^u(a_i,b_i,\rho_i)(\Pr(a,b_i,\rho_i,b_j,\rho_j) - \Pr(a,b_i,\rho_i)),$$

and $\Pr(s|a,b_i,\rho_i,b_j,\rho_j) = \sum_{t_j \in \rho_j^{-1}(s_j)} \Pr(s_{-j},t_j|a_{-j},b_j,b_i,\rho_i)$. Secondly, notice that

$$\begin{aligned}
\forall i \in I, a_i \in B_i, \qquad & \sum_{a_{-i}} \mu^m(a)(\Delta v_i(a,\lambda_i) - z\,\|\Delta \Pr(a,\lambda_i)\|) \;\le \\
& \mu_i^m(a_i) \sum_{a_{-i}} \mu_{-i}^m(a_{-i})(\Delta v_i(a_i,\lambda_i)^* - z\,\|\Delta \Pr(a,\lambda_i)\|) \;\le \\
& \mu_i^m(a_i) \sum_{a_{-i}} \mu_{-i}^0(a_{-i})(\Delta v_i(a_i,\lambda_i)^* - z\,\|\Delta \Pr(a,\lambda_i)\|) \;\le \\
& \qquad \sum_{a_{-i}} \mu^0(a)(\Delta v_i(a_i,\lambda_i)^* - z\,\|\Delta \Pr(a,\lambda_i)\|).
\end{aligned}$$

Indeed, the first inequality is obvious. The second one follows by repeated application of the previously derived inequality $\|\Delta \Pr(a,\lambda_i)\| \le \|\Delta \Pr(a,\widehat{\lambda}_j^u,\lambda_i)\|$ for each agent $j \neq i$ separately $m$ times. The third inequality follows because (i) $\mu_i^m(a_i) \ge \mu_i^0(a_i)$ for all $m$ and $a_i \in B_i$, since $B_i$ is a $\widehat{\lambda}_i^u$-absorbing set, and (ii) $\sum_{a_{-i}} \Delta v_i(a_i,\lambda_i)^* - z\,\|\Delta \Pr(a,\lambda_i)\| \le 0$ for every $(i,a_i)$ by Lemma B.6. Therefore, $W_d(\mu^\infty,\lambda) \le W_d(\mu^m,\lambda) \le W_d(\mu^0,\lambda) \le C$. Thirdly,

$$\begin{aligned}
\forall i \in I, a_i \notin B_i, \qquad & \sum_{a_{-i}} \mu_{-i}^m(a_{-i})(\Delta v_i(a,\lambda_i^e) - z\,\|\Delta \Pr(a,\lambda_i^e)\|) \;\le \\
& \sum_{a_{-i}} \mu_{-i}^m(a_{-i})(\Delta v_i(a_i,\lambda_i^e)^* - z\,\|\Delta \Pr(a,\lambda_i^e)\|) \;\le \\
& \sum_{a_{-i}} \mu_{-i}^0(a_{-i})(\Delta v_i(a_i,\lambda_i^e)^* - z\,\|\Delta \Pr(a,\lambda_i^e)\|) \;\le\; 0.
\end{aligned}$$

The first inequality is again obvious, the second inequality follows by repeated application of $\|\Delta \Pr(a,\lambda_i)\| \le \|\Delta \Pr(a,\widehat{\lambda}_j^u,\lambda_i)\|$, and the third one follows from Lemma B.9. Hence, $W_e(\mu^m,\lambda) \le 0$ for every $m$, therefore $W_e(\mu^\infty,\lambda) \le 0$. This completes the proof. (This proof extends Nau and McCardle (1990) and Myerson (1997) by including transfers.) $\square$

**Lemma B.11.** *The conditions of Theorem 8 imply that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\sum_i \Delta v_i(\sigma, \lambda_i) \geq \varepsilon$ implies that $\sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\| \geq \delta$ for some $\mu \in \Delta(A)$ with $\sum_i \Delta v_i(\mu, \lambda_i) \leq \overline{z} \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\|$.*

*Proof.* Otherwise, there exists $\varepsilon > 0$ such that for every $\delta > 0$ some $\lambda^\delta$ exists with $\sum_i \Delta v_i(\sigma, \lambda_i^\delta) \geq \varepsilon$ but $\sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\| < \delta$ whenever $\mu \in \Delta(A)$ satisfies the given inequality $\sum_i \Delta v_i(\mu, \lambda_i) \leq \overline{z} \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\|$. If $\lambda^\delta$ is bounded for every $\delta$ then $\{\lambda^\delta\}$ has a convergent subsequence with limit $\lambda^0$. But this $\lambda^0$ violates the conditions of Theorem 8, so assume that $\{\lambda^\delta\}$ is unbounded. A deviation plan $\alpha_i^r$ is called *relatively undetectable* if $\sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\| = 0$ whenever $\mu \in \Delta(A)$ satisfies $\sum_i \Delta v_i(\mu, \lambda_i) \leq \overline{z} \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\|$. Call $\mathscr{D}_i^r$ the set of relatively undetectable plans. A deviation plan $\alpha_i^s$ is called *relatively detectable* if

$$\max_{(p, \alpha_i, \alpha_i^r)} \{p : p\alpha_i^r + (1-p)\alpha_i = \alpha_i^s, \alpha_i \in \mathscr{D}_i, \alpha_i^r \in \mathscr{D}_i^r, p \in [0,1]\} \;=\; 0.$$

Let $\mathscr{D}_i^s$ be the set of relatively detectable plans. By the same argument as for Lemma B.7, $\mathscr{D}_i^s$ is a compact set, therefore, by the same argument as for Lemma B.8,

$$\gamma_i^s := \min_{\alpha_i^s \in \mathscr{D}_i^s} \max_{\mu \in \Delta(A)} \left\{ \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \alpha_i^s)\| : \sum_{i \in I} \Delta v_i(\mu, \lambda_i) \leq \overline{z} \sum_{(i,a)} \mu(a) \|\Delta \Pr(a, \lambda_i)\| \right\} > 0.$$

Without loss, $\lambda_i^\delta = \lambda_i^{r,\delta} + \lambda_i^{s,\delta}$, where $\lambda_i^{r,\delta}$ is relatively undetectable and $\lambda_i^{s,\delta}$ is relatively detectable. By assumption, $\lambda_i^{r,\delta}$ is $\sigma$-unprofitable, so $\sum_{(b_i,\rho_i)} \lambda_i^{s,\delta}(a_i, b_i, \rho_i)$ is bounded below by $\beta > 0$, say. (Otherwise, $\sum_i \Delta v_i(\sigma, \lambda_i^\delta) < \varepsilon$ for small $\delta > 0$.) But this implies that

$$\max_{\mu \in \Delta(A)} \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^\delta) \right\| \;=\; \max_{\mu \in \Delta(A)} \sum_{(i,a)} \mu(a) \left\| \Delta \Pr(a, \lambda_i^{s,\delta}) \right\| \;\geq\; \beta \gamma_i^s \;>\; 0.$$

But this contradicts our initial assumption, which establishes the result. □

# References


Alchian, A. and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777–795. 1

Aoyagi, M. (2005): "Collusion Through Mediated Communication in Repeated Games with Imperfect Private Monitoring," *Economic Theory*, 25, 455–475. 22

Che, Y.-K. and J. Kim (2006): "Robustly Collusion-Proof Implementation," *Econometrica*, 74, 1063–1107. 29

Cheslow, D. (2005): "Students, Bartenders Cited in Liquor Sting," *The Daily Northwestern,* http://www.dailynorthwestern.com/. 1

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66, 597–626. 14

Cremer, J. and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257. 29

d'Aspremont, C., J. Cremer, and L.-A. Gérard-Varet (2004): "Balanced Bayesian Mechanisms," *Journal of Economic Theory*, 115, 385–396. 19, 22

d'Aspremont, C. and L.-A. Gérard-Varet (1998): "Linear Inequality Methods to Enforce Partnerships under Uncertainty: An Overview," *Games and Economic Behavior*, 25, 311–336. 14, 18, 21, 28

Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997–1039. 2, 14, 18, 22

Holmström, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340. 4, 6, 15

Kandori, M. (2003): "Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Monitoring," *Econometrica*, 71, 345–353. 14, 22

Kandori, M. and H. Matsushima (1998): "Private Observation, Communication, and Collusion," *Econometrica*, 66, 627–652. 14, 22

Kandori, M. and I. Obara (2006): "Efficiency in Repeated Games Revisited: The Role of Private Strategies," *Econometrica*, 74, 499–519. 22

Kosenok, G. and S. Severinov (forthcoming): "Individually Rational, Balanced-Budget Bayesian Mechanisms and the Allocation of Surplus," *Journal of Economic Theory.* 22

Legros, P. and H. Matsushima (1991): "Efficiency in Partnerships," *Journal of Economic Theory*, 55, 296–322. 14, 21, 28

Legros, P. and S. Matthews (1993): "Efficient and Nearly Efficient Partnerships," *Review of Economic Studies*, 60, 599–611. 14, 21, 28

Myerson, R. (1986): "Multistage Games with Communication," *Econometrica*, 54, 323–358. 2

——— (1997): "Dual Reduction and Elementary Games," *Games and Economic Behavior*, 21, 183–202. 43

Nau, R. F. and K. F. McCardle (1990): "Coherent Behavior in Noncooperative Games," *Journal of Economic Theory*, 50, 424–444. 43

Radner, R., R. Myerson, and E. Maskin (1986): "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies*, 53, 59–69. 5

Rahman, D. (2005a): "Optimum Contracts with Public and Private Monitoring," Mimeo. 0

——— (2005b): "Team Formation and Organization," Ph.D. dissertation, UCLA. 0

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press. 12

Tomala, T. (2005): "On Subgame-Perfect Communication Equilibria in Repeated Games with Imperfect Monitoring," Mimeo. 15, 22

TSA (2004): "Guidance on Screening Partnership Program," *Transportation Security Administration*, http://www.tsa.gov/assets/pdf/ SPP_OptOut_Guidance_6.21.04.pdf. 1