B Ancillary Results

Lemma B.1. *Let* $\Pr(a_i, s_i)$ *be the vector defined pointwise by* $\Pr(a_i, s_i)(a_{-i}, s_{-i}) = \Pr(s|a)$ *for each* (a_{-i}, s_{-i}). *Every disobedience is detectable if* $\Pr$ *exhibits* conic independence, *i.e.,*

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

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. Otherwise, there exists σ_i such that $\sigma_i(b_i, \rho_i|a_i) > 0$ for some $a_i \neq b_i$ and

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

Write $\lambda_i(a_i, s_i, b_i, t_i) := \sum_{\{\rho_i:\rho_i(t_i)=s_i\}} \sigma_i(b_i, \rho_i|a_i)$. 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 $\Pr(s|a) = \sum_{(b_i,t_i)} \lambda_i(a_i, s_i, b_i, t_i) \Pr(s_{-i}, t_i|a_{-i}, b_i)$ for all (i, a, s). Without loss, $\lambda_i(a_i, s_i, a_i, s_i) = 0$ for some (a_i, s_i). To see this, note first that $\lambda_i(a_i, s_i, a_i, s_i) = 1$ for all (a_i, s_i) is impossible because $\sigma_i \geq 0$ is assumed disobedient. If $\lambda_i(a_i, s_i, a_i, s_i) \neq 1$, subtract $\lambda_i(a_i, s_i, a_i, s_i) \Pr(s|a)$ from both sides and divide by $1 - \lambda_i(a_i, s_i, a_i, s_i)$. Now $\Pr(a_i, s_i) \in \text{cone}\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\}$ for some (a_i, s_i).

Proof of Theorem 8. By Lemma B.1, detectability of every disobedience is implied by conic independence. In turn, this is implied by *linear* independence, or full row rank, for all i, of the $|A_i|\,|S_i| \times |A_{-i}|\,|S_{-i}|$ matrix with 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 when $|A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}|$ 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 $|A_i| = |A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}| - (|A_{-i}| - 1) = |A_{-i}|\,(|S_{-i}| - 1) + 1$ for all i. Similarly, there are $|A_i| - 1$ redundant rows when $|S_{-i}| = 1$. It remains to show that $|A_i| - 1 \leq |A_{-i}|\,(|S_{-i}| - 1)$ follows from $|A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}|$ if both $|S_i| > 1$ and $|S_{-i}| > 1$. The latter inequality implies $2\,|A_i| \leq |A_{-i}|\,|S_{-i}|$ if $|S_i| > 1$, so $|A_i| \leq |A_{-i}|\,|S_{-i}|\,/2$. This implies $|A_i| \leq |A_{-i}|(|S_{-i}|-1)$ if $|S_{-i}| > 1$, so $|A_i| - 1 \leq |A_{-i}|(|S_{-i}|-1)$. Since the intersection of finitely many generic sets (one per agent) is generic, the result follows.

Let $\mathscr{D}_i = \Delta(A_i \times R_i)^{A_i}$ be the space of strategies σ_i for a agent i and $\mathscr{D} = \prod_i \mathscr{D}_i$ the set of strategy profiles $\sigma = (\sigma_1, \ldots, \sigma_n)$. Call μ *enforceable within* some vector $z \in \mathbb{R}^I_+$ if there is a scheme ξ that satisfies $(*)$ and $-\mu(a)z_i \leq \xi_i(a,s) \leq \mu(a)z_i$ for all (i,a,s). Next, we provide a lower bound on z so that μ is enforceable within z.

Lemma B.2. (i) *A correlated strategy μ is enforceable within $z \in \mathbb{R}^I_+$ if and only if*

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

(ii) *If μ is enforceable then $V_\mu(z) = 0$ for some $z \in \mathbb{R}^I_+$. If not then* $\sup_z V_\mu(z) > 0$.

(iii) *A correlated strategy μ is enforceable if and only if $\overline{z}_i < +\infty$ for every agent i, where*

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

and, whenever $\mathscr{F}_i = \emptyset$, $\overline{z}_i := +\infty$ exactly when $\max_{\sigma_i} \Delta v_i(\mu, \sigma_i) > 0$.[38]

(iv) If $\overline{z}_i < +\infty$ for every i then $V_\mu(z) = 0$ if and only if $z_i \geq \overline{z}_i$ for all i.

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

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

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

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 ξ adapted to μ (i.e., such that $\xi_i(a,s) = 0$ whenever $\mu(a) = 0$) that minimizes the burden $\varepsilon_i(a_i)$ of relaxing incentive constraints. By construction, μ is enforceable with transfers bounded by z if and only if there is a feasible ξ with $\varepsilon_i(a_i) = 0$ for all (i, a_i), i.e., the value of the problem is zero. Since μ is assumed enforceable, such z exists. The dual of this problem is:

$$\min_{\sigma, \beta \geq 0} \sum_{(i,a)} \mu(a) [z_i \sum_{s \in S} \mu(a)(\beta_i^+(a,s) + \beta_i^-(a,s)) - \Delta v_i(a, \sigma_i)] \;\text{ s.t.} \sum_{(b_i, \rho_i)} \sigma_i(b_i, \rho_i | a_i) \leq 1 \;\forall (i, a_i),$$

$$\Delta \Pr(s | a, \sigma_i) \;=\; \beta_i^+(a,s) - \beta_i^-(a,s) \quad \forall i \in I, a \in \operatorname{supp} \mu, s \in S.$$

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

Since $\beta_i^{\pm}(a,s) \geq 0$, it is not difficult to see that both $\beta_i^+(a,s) = \max\{\Delta \Pr(s|a,\sigma_i), 0\}$ and $\beta_i^-(a,s) = \min\{\Delta \Pr(s|a,\sigma_i), 0\}$. Therefore, $\beta_i^+(a,s) + \beta_i^+(a,s) = |\Delta \Pr(s|a,\sigma_i)|$. Furthermore, $\|\Delta \Pr(a,\sigma_i)\| = \sum_s |\Delta \Pr(s|a,\sigma_i)|$, so the dual is now equivalent to

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

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

To prove (ii), the first sentence is obvious. The second follows by Theorem 1: if μ is not enforceable then a μ-profitable, supp μ-undetectable plan σ_i exists, so $V_\mu(z) > 0$ for all z.

For (iii), if μ is not enforceable then there is a μ-profitable, supp μ-undetectable deviation plan σ_i^*. Approaching σ_i^* from $\mathscr{F}_i$ (e.g., with mixtures of σ_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 μ is enforceable. If the sup defining $\overline{z}_i$ is attained, we are done. If not, it is approximated by a sequence of supp μ-detectable deviation plans that converge to a supp μ-undetectable one. Since μ is enforceable, the limit is unprofitable. Let

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

Since every μ-profitable deviation plan is detectable by Theorem 1, it follows that $F_i^\mu(\delta) > 0$ for all $\delta > 0$, and $\overline{z}_i = (\lim_{\delta \downarrow 0} F_i^\mu(\delta)/\delta)^{-1}$. Hence, it suffices to show $\lim_{\delta \downarrow 0} F_i^\mu(\delta)/\delta > 0$. To this end, by adding variables like β above, the dual problem for F_i^μ is equivalent to:

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

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

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

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

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

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

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

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

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

Proof. The first family of primal constraints require ξ to be adapted to $\eta + (1-p)\mu$, so for any z, (η, p, ξ) solves the primal if and only if $\eta + (1-p)\mu$ is exactly enforceable with ξ. (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 $z \geq 0$.*

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

$F_\mu^\varepsilon(z) \to -\infty$ as $z \to \infty$ for some $\varepsilon > 0$ if and only if μ is virtually enforceable.

Proof. The dual of the problem defining $F^{\varepsilon}_{\mu}(z)$ is

$$F^{\varepsilon}_{\mu}(z) \;=\; \min_{\delta,\eta\geq 0,x} \; -\delta\varepsilon \;\; \text{s.t.} \;\; \sum_{a\in A}\eta(a) \;=\; 1,$$

$$\forall(i,a,s), \qquad -(\eta(a)+\delta\mu(a))z \;\leq\; x_i(a,s) \;\leq\; (\eta(a)+\delta\mu(a))z,$$

$$\forall(i,a_i,b_i,\rho_i), \qquad \sum_{a_{-i}}(\eta(a)+\delta\mu(a))\Delta v_i(a,b_i) \;\leq\; \sum_{a_{-i}} x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i).$$

Since clearly $\varepsilon > 0$ does not affect the dual feasible set, if $F^{\varepsilon}_{\mu}(z) \to -\infty$ 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, $F^{\varepsilon}_{\mu}(z) \to -\infty$ for every $\varepsilon > 0$. If $V_\mu(z) = 0$ for some z we are done by monotonicity of V_μ. Otherwise, suppose that $V_\mu(z) > 0$ for all $z > 0$. Let (λ,κ) be an optimal dual solution for $V_\mu(z)$ in Lemma B.3. By optimality, $\kappa = \min_{\eta\in\Delta(A)}\sum_i \Delta v_i(\eta,\lambda_i) - z\sum_{(i,a)}\eta(a)\,\|\Delta\Pr(a,\lambda_i)\|$. Therefore, by the second dual constraint in $V_\mu(z)$ of Lemma B.3,

$$V_\mu(z) \;=\; 1+\kappa \;=\; 1+F^{V_\mu(z)}_{\mu}(z) \;=\; 1-\delta V_\mu(z),$$

where δ is an optimal solution to the dual with $\varepsilon = V_\mu(z)$. Rearranging, $V_\mu(z) = 1/(1+\delta)$. Finally, $F^{\varepsilon}_{\mu}(z) \to -\infty$ as $z\to\infty$ if and only if $\delta\to\infty$, if and only if $V_\mu(z)\to 0$.

Lemma B.5. *Fix any $\varepsilon > 0$ and let $B = \operatorname{supp}\mu$. If every B-disobedience is detectable then for every $C \leq 0$ there exists $z \geq 0$ such that $G_\mu(z) \leq C$, where*

$$\Delta v_i(a_i)^* := \max_{(a_{-i},b_i)}\{\Delta v_i(a,b_i)\}, \qquad \Delta v_i(a_i,\lambda_i)^* \;:=\; \Delta v_i(a_i)^* \sum_{(a_i,b_i\neq a_i,\rho_i)} \lambda_i(a_i,b_i,\rho_i), \;\; \textit{and}$$

$$G_\mu(z) \;:=\; \max_{\lambda\geq 0}\sum_{(i,a)}\|\Delta v_i(a_i,\lambda_i)\| - z\sum_{(i,a)}\|\Delta\Pr(a,\lambda_i)\| \qquad \text{s.t.}$$

$$\forall i\in I, a_i\notin B_i,\; \lambda_i(a_i) \;=\; 0, \qquad \textit{and} \qquad \sum_{i\in I}\Delta v_i(\mu,\lambda_i) - z\sum_{(i,a)}\mu(a)\,\|\Delta\Pr(a,\lambda_i)\| \;\geq\; \varepsilon.$$

Proof. The dual of this problem is given by

$$G_\mu(z) \;=\; \min_{\delta\geq 0,x} \; -\delta\varepsilon \;\; \text{s.t.}$$

$$\forall(i,a,s), \qquad -(1+\delta\mu(a))z \;\leq\; x_i(a,s) \;\leq\; (1+\delta\mu(a))z,$$

$$\forall(i,a_i\in B_i,b_i,\rho_i), \qquad \sum_{a_{-i}}\delta\mu(a)\Delta v_i(a,b_i) + \mathbf{1}_{\{a_i\neq b_i\}}\Delta v_i(a_i)^* \;\leq\; \sum_{a_{-i}} x_i(a)\cdot\Delta\Pr(a,b_i,\rho_i),$$

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 $F_\mu^\varepsilon(z)$ except that the incentive constraints are only indexed by $a_i \in B_i$. Now, every B-disobedience is detectable if and only if there is an incentive scheme x such that

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

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

Lemma B.6. *If every B-disobedience is detectable 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, $a_i \in B_i$, let $\mu(a) = 1/|A_{-i}|$ for all a_{-i} in the proof of Lemma B.2 (iii).

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

Lemma B.7. *The set $\mathscr{D}^e = \{\sigma \in \mathscr{E} : \forall (i, a_i), \sum_{(b_i, \rho_i)} \sigma_i(a_i, b_i, \rho_i) = 1\}$ 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 $\{\sigma^m\} \subset \mathscr{D}^e$ such that $\sigma^m \to \sigma^*$. For any $\sigma \in \mathscr{D}$, let

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

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

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

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

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 $\{\sigma_i^{e,m}\} \subset \mathscr{D}^e$ with $\|\Delta \Pr(a, \sigma_i^{e,m})\| \to 0$ for all (i, a) as $m \to \infty$, hence $\gamma > 0$.

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

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

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

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

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

Lemma B.10. *Fix any $\varepsilon > 0$. If every B-disobedience is detectable then for every $C \leq 0$ there exists $z \geq 0$ such that for every $\lambda \geq 0$ with*

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

there exists $\eta \in \Delta(A)$ such that

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

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

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

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

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

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

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

Iterate this rule to obtain a sequence $\{\eta_i^m\}$ with limit $\eta_i^\infty \in \Delta(A_i)$. By construction, η_i^∞ is a λ_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.,

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

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

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

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

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

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

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

The first inequality is again obvious, the second inequality follows by repeated application of $\|\Delta \Pr(a, \lambda_i)\| \leq \|\Delta \Pr(a, \widehat{\lambda}_j^u, \lambda_i)\|$, and the third one follows from Lemma B.9. Hence, $W_e(\eta^m, \lambda) \leq 0$ for every m, therefore $W_e(\eta^\infty, \lambda) \leq 0$. This completes the proof.

Lemma B.11. *The conditions of Theorem 4 imply that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\sum_i \Delta v_i(\mu, \lambda_i) \geq \varepsilon$ implies that $\sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \lambda_i)\| \geq \delta$ for some $\eta \in \Delta(A)$ with $\sum_i \Delta v_i(\eta, \lambda_i) \leq \overline{z} \sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \lambda_i)\|$.*

Proof. Otherwise, there exists $\varepsilon > 0$ such that for every $\delta > 0$ some λ^δ exists with $\sum_i \Delta v_i(\mu, \lambda_i^\delta) \geq \varepsilon$, but $\sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \lambda_i)\| < \delta$ whenever $\eta \in \Delta(A)$ satisfies the given inequality $\sum_i \Delta v_i(\eta, \lambda_i) \leq \overline{z} \sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \lambda_i)\|$. If $\{\lambda^\delta\}$ is uniformly bounded then it has a convergent subsequence with limit λ^0. But this λ^0 violates the conditions of Theorem 4, so $\{\lambda^\delta\}$ must be unbounded. Call a deviation σ_i^r *relatively undetectable* if given $\eta \in \Delta(A)$, $\sum_i \Delta v_i(\eta, \sigma_i^r) \leq \overline{z} \sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \sigma_i^r)\|$ implies $\sum_{(i,a)} \eta(a) \|\Delta \Pr(a, \sigma_i^r)\| = 0$. Call $\mathscr{D}_i^r$ the set of relatively undetectable plans. A deviation σ_i^s is called *relatively detectable* if

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

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,

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

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 μ-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, $\sum_i \Delta v_i(\mu, \lambda_i^\delta) < \varepsilon$ for small $\delta > 0$.) But this implies that

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

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

Proof of Theorem 4. For sufficiency, suppose that μ is virtually enforceable, so there is a sequence $\{\mu^m\}$ such that μ^m is enforceable for every m and $\mu^m \to \mu$. Without loss, assume that supp $\mu^m \supset$ supp μ for all m. If $\mu^m = \mu$ for all large m then μ is enforceable and the condition of Theorem 4 is fulfilled with $\eta = \mu$, so suppose not. If there exists m and m' such that $\mu^m = p\mu^{m'} + (1-p)\mu$ then incentive compatibility with respect to m yields that $\sum_{a_{-i}} \mu^m(a)\Delta v_i(a,\sigma_i) \leq \sum_{a_{-i}} \mu^m(a)\zeta_i^m(a) \cdot \Delta \Pr(a,\sigma_i) \leq \sum_{a_{-i}} \mu^m(a)\overline{z} \left\|\Delta \Pr(a,\sigma_i)\right\|$ for every σ_i, where $\overline{z} = \max_{(i,a,s)} |\zeta_i^m(a,s)| + 1$ and ζ^m enforces μ^m for each m. For large m', $\mu^{m'}$ is sufficiently close to μ that if σ_i is μ-profitable then $\sum_{a_{-i}} \mu^{m'}(a)\Delta v_i(a,\sigma_i) > 0$, so σ_i is detectable. Therefore, $\sum_{a_{-i}} \mu^m(a)\Delta v_i(a,\sigma_i) < \sum_{a_{-i}} \mu^m(a)\overline{z} \left\|\Delta \Pr(a,\sigma_i)\right\|$.

If no m and m_1 exist with $\mu^m = p\mu^{m_1} + (1-p)\mu$ then μ^{m_2} exists such that its distance from μ is less than the positive minimum distance between μ and the affine hull of $\{\mu^m, \mu^{m_1}\}$. Therefore, the lines generated by μ^m and μ^{m_1} and μ^{m_1} and μ^{m_2} are not collinear. Proceeding inductively, pick $C = \{\mu^{m_1}, \ldots, \mu^{m_{|A|}}\}$ such that its affine space is full-dimensional in $\Delta(A)$. Since we are assuming that μ is not enforceable, it lies outside conv C. Let $\widehat{\mu} = \sum_k \mu^{m_k} / |A|$ and $B_\varepsilon(\widehat{\mu})$ be the open ε-ball around $\widehat{\mu}$ for some $\varepsilon > 0$. By construction, $B_\varepsilon(\widehat{\mu}) \subset \text{conv } C$ for $\varepsilon > 0$ sufficiently small, so there exists $\widehat{\mu}' \in B_\varepsilon(\widehat{\mu})$ such that $p\widehat{\mu} + (1-p)\mu = \widehat{\mu}'$ for some p such that $0 < p < 1$. Now, by the previous paragraph, the condition of Theorem 4 holds.

For necessity, if μ is not virtually enforceable then $1 \geq V_\mu(z) \geq C > 0$ for every z, where V_μ is defined in Lemma B.3. Let (λ^z, κ^z) solve $V_\mu(z)$ for every z. Given $\eta \in \Delta(A)$,

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

By the condition of Theorem 4, $\overline{z}$ exists with $\sum_{(i,a)} \Delta v_i(\eta^z, \lambda_i^z) < \overline{z} \sum_{(i,a)} \eta^z(a) \left\|\Delta \Pr(a, \lambda_i^z)\right\|$ and $\sum_{(i,a)} \eta^z(a) \left\|\Delta \Pr(a, \lambda_i^z)\right\| > 0$ for some η^z, since λ_i^z is μ-profitable for some i. Hence, $C \leq 1 + (\overline{z} - z) \sum_{(i,a)} \eta^z(a) \left\|\Delta \Pr(a, \lambda_i^z)\right\|$, i.e., $z - \overline{z} \leq (1-C) / \sum_{(i,a)} \eta^z(a) \left\|\Delta \Pr(a, \lambda_i^z)\right\|$. This inequality must hold for every z, therefore $\sum_{(i,a)} \eta^z(a) \left\|\Delta \Pr(a, \lambda_i^z)\right\| \to 0$ as $z \to \infty$. But this contradicts Lemma B.11, since $\sum_i \Delta v_i(\mu, \lambda_i^z) \geq C$, completing the proof.