Econometrica Supplementary Material

SUPPLEMENT TO "MECHANISMS WITH EVIDENCE: COMMITMENT AND ROBUSTNESS"

(*Econometrica*, Vol. 87, No. 2, March 2019, 529–566)

ELCHANAN BEN-PORATH
Department of Economics and Federmann Center for the Study of Rationality, Hebrew University

EDDIE DEKEL
Economics Department, Northwestern University and School of Economics, Tel Aviv University

BARTON L. LIPMAN
Department of Economics, Boston University

SA. EXAMPLE FOR ROBUSTNESS DEFINITION

IN THIS SECTION, we give an example to show that in games with evidence, even with independent private values, the requirements of robust incentive compatibility, dominant strategy incentive compatibility, and ex post incentive compatibility differ.

The reason that ex post incentive compatibility and robust incentive compatibility are not equivalent is that robust incentive compatibility requires truth-telling to be optimal even when other agents deviate from truth-telling with maximal evidence. In the absence of evidence, the fact that we have independent private values means that agent i is unaffected by whether the claims of other agents are true or not. Hence these two notions would be the same in that case. But with evidence, we can have reports by the other agents that would be impossible under truth-telling with maximal evidence.

The reason that dominant strategy incentive compatibility is not the same as robust incentive compatibility is that dominant strategy incentive compatibility only requires that truth-telling and maximal evidence be a best reply to any strategy function by the other agents. In the absence of evidence, the other agents could be playing constant strategies, implying that truth-telling and maximal evidence must be a best reply to any reports by the other agents. In mechanisms with evidence, however, constant strategies may not be possible.

To see both points in a simple example, suppose $I = 2$ and $T_i = \{\alpha_i, \beta_i\}$, $i = 1, 2$. Suppose $\mathcal{E}_i(\alpha_i) = \{\{\alpha_i\}\}$ and $\mathcal{E}_i(\beta_i) = \{\{\beta_i\}\}$, $i = 1, 2$. Suppose the principal has just two actions, denoted 0 and 1. Assume $u_1(a) = a$ and $u_2(a) = 0$ for all a.[1] Say that agent i reports consistently if she reports $(\alpha_i, \{\alpha_i\})$ or $(\beta_i, \{\beta_i\})$ and reports inconsistently otherwise. Note that all three versions of incentive compatibility say that consistent reports are optimal and differ only in the circumstances under which consistent reports are required to be optimal. Assume that the prior probability that $t_2 = \beta_2$ is strictly below 1/2.

Consider the mechanism where the principal chooses $a = 1$ if one of the following is true. First, 1's report is consistent and 2's report (consistent or not) has evidence $\{\alpha_2\}$.

Elchanan Ben-Porath: benporat@math.huji.ac.il
Eddie Dekel: dekel@northwestern.edu
Barton L. Lipman: blipman@bu.edu

We thank numerous seminar audiences and Joel Sobel and four anonymous referees for useful comments and suggestions. We also thank the National Science Foundation, Grant SES-0820333 (Dekel), and the US–Israel Binational Science Foundation for support for this research.

[1] It is not difficult to give a symmetric but more complex example where neither agent is completely indifferent.

 https://doi.org/10.3982/ECTA14991

Second, both reports are consistent and 2's evidence presentation is $\{\beta_2\}$. Third, both reports are inconsistent and 2's evidence presentation is $\{\beta_2\}$. If the reports do not satisfy one of these three conditions, then the principal chooses $a=0$. Because 2 is indifferent between $a=0$ and $a=1$, the mechanism satisfies robust incentive compatibility for him. However, it is not robustly incentive compatible for 1. To see this, simply note that 1's best response to a report by 2 of $(\alpha_2, \{\beta_2\})$ is to be inconsistent.

On the other hand, this mechanism is both ex post incentive compatible and dominant strategy incentive compatible for 1. To see that it is ex post incentive compatible, note that if 2 is consistent, then 1's best response is always to be consistent, regardless of the type profile. To see that it is dominant strategy incentive compatible, note that for any feasible strategy for 2, 1's expected payoff to any consistent report is at least the probability that $t_2=\alpha_2$, while the payoff to any inconsistent report is at most the probability that $t_2=\beta_2$. Since the former strictly exceeds the latter, reporting consistently is a dominant strategy.

SB. EQUILIBRIUM DEFINITION

Our definition of perfect Bayesian equilibrium is identical to that of Fudenberg and Tirole (1991) but adapted to allow type-dependent sets of feasible actions.

Given $\sigma_{-i} \in \Sigma_{-i}$, $\sigma_P \in \Sigma_P$, $a \in A$, and $(s_i, e_i) \in T_i \times \mathcal{E}_i$, let

$$Q_i(a \mid s_i, e_i, \sigma_{-i}, \sigma_P) = \mathrm{E}_{t_{-i}} \sum_{(s_{-i}, e_{-i})} \sigma_P(a \mid s, e) \prod_{j \neq i} \sigma_j(s_j, e_j \mid t_j).$$

This is the probability the principal chooses allocation a given that she uses strategy σ_P, agents other than i use strategies σ_j, $j \neq i$, and agent i reports s_i and presents evidence e_i. Let $\mu : T \times \mathcal{E} \to \Delta(T)$ be the function that specifies the beliefs of the principal.

We say that $(\sigma_1, \dots, \sigma_I, \sigma_P, \mu)$ is a perfect Bayesian equilibrium if the following conditions hold. First, for every i and every $t_i \in T_i$, $\sigma_i(s_i, e_i \mid t_i) > 0$ implies

$$(s_i, e_i) \in \arg\max_{s_i' \in T_i, e_i' \in \mathcal{E}_i(t_i)} \sum_{a \in A} Q_i\big(a \mid s_i', e_i', \sigma_{-i}, \sigma_P\big) u_i(a, t_i).$$

(Recall that $Q_i(a \mid s_i, e_i, \sigma_{-i}, \sigma_P)$ is the probability that the actions of the other agents are such that the principal chooses action a given that i uses action (s_i, e_i) and given strategies σ_{-i} and σ_P.) Second, for every $(s, e) \in T \times \mathcal{E}$, $\sigma_P(a \mid s, e) > 0$ implies

$$a \in \arg\max_{a' \in A} \sum_{t \in T} \mu(t \mid s, e) v\big(a', t\big).$$

Third, for every (s, e), $\mu(\cdot \mid s, e)$ respects independence across agents. That is, i's report (s_i, e_i) only affects the principal's beliefs about t_i and his beliefs about t_i and t_j respect independence for all $i \neq j$. Formally, we have functions $\mu_i : T_i \times \mathcal{E}_i \to \Delta(T_i)$ such that, for all $t \in T$ and all $(s, e) \in T \times \mathcal{E}$,

$$\mu(t \mid s, e) = \prod_i \mu_i(t_i \mid s_i, e_i).$$

Fourth, for all (s, e), $\mu(\cdot \mid s, e)$ respects feasibility. That is, the principal's beliefs must put zero probability on any type which is infeasible given (s, e). Formally, for every $t_i \in T_i$ and $(s_i, e_i) \in T_i \times \mathcal{E}_i$, we have $\mu_i(t_i \mid s_i, e_i) = 0$ if $e_i \notin \mathcal{E}_i(t_i)$.

Finally, the principal's beliefs are consistent with Bayes's rule whenever possible in the sense that for every $(s_i, e_i) \in T_i \times \mathcal{E}_i$ such that there exists t_i with $\sigma_i(s_i, e_i \mid t_i) > 0$, we have

$$\mu_i(t_i \mid s_i, e_i) = \frac{\sigma_i(s_i, e_i \mid t_i)\rho_i(t_i)}{\sum_{t_i' \in T_i} \sigma_i(s_i, e_i \mid t_i')\rho_i(t_i')}.$$

(Recall that ρ_i is the principal's prior over t_i.)

SC. PROOF OF THEOREM 2

Recall that T_i^0 is the set of t_i such that $\mathcal{E}_i(t_i) = \{T_i\}$. We first show there exists v_i^* solving

$$v_i^* = \mathrm{E}_{t_i}\big[v_i(t_i) \mid t_i \in T_i^0 \text{ or } v_i(t_i) \le v_i^*\big]. \tag{S1}$$

If $T_i = T_i^0$, then $v_i^* = \mathrm{E}_{t_i}(v_i(t_i))$ satisfies (S1). If $T_i^0 = \emptyset$, then $v_i^* = \min_{t_i \in T_i} v_i(t_i)$ satisfies (S1). In what follows, assume $T_i^0 \neq \emptyset$ and $T_i^0 \neq T_i$.

Write $T_i \setminus T_i^0 = \{t_i^1, \dots, t_i^N\}$ where $v_i(t_i^n) < v_i(t_i^{n+1})$. (If we have $t_i, t_i' \in T_i \setminus T_i^0$ with $t_i \neq t_i'$ and $v_i(t_i) = v_i(t_i')$, we can treat these two types as one for this calculation.) For $n = 1, \dots, N$, let

$$g_i^n = \mathrm{E}_{t_i}\big[v_i(t_i) \mid t_i \in T_i^0 \text{ or } t_i = t_i^k, \text{ for some } k \le n\big]$$

and let $g_i^0 = \mathrm{E}_{t_i}(v_i(t_i) \mid t_i \in T_i^0)$.

Suppose there is no solution to equation (S1). If $g_i^0 \le v_i(t_i^1)$, then $v_i^* = g_i^0$ satisfies (S1). Hence $g_i^0 > v_i(t_i^1)$. But g_i^1 is a convex combination of $v_i(t_i^1)$ and g_i^0, with strictly positive weight on each term, so $v_i(t_i^1) < g_i^1 < g_i^0$. Again, if $g_i^1 \le v_i(t_i^2)$, then $v_i^* = g_i^1$ satisfies (S1), so we must have $g_i^1 > v_i(t_i^2)$, implying $v_i(t_i^2) < g_i^2 < g_i^1$. Similar reasoning gives $g_i^{n-1} > g_i^n > v_i(t_i^n)$ for $n = 1, \dots, N$. In particular, $g_i^N > v_i(t_i^N)$. But $g_i^N = \mathrm{E}_{t_i}[v_i(t_i)]$, so $v_i^* = g_i^N$ solves equation (S1), a contradiction. Hence a solution exists.

To show uniqueness, suppose v_i^1 and v_i^2 both solve (S1) where $v_i^1 > v_i^2$. For $k = 1, 2$, let

$$T_i^k = T_i^0 \cup \big\{t_i \in T_i \setminus T_i^0 \mid v_i(t_i) \le v_i^k\big\},$$

so $v_i^k = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in T_i^k]$. Since $v_i^1 > v_i^2$, we have $T_i^2 \subset T_i^1$ and

$$T_i^1 \setminus T_i^2 = \big\{t_i \in T_i \setminus T_i^0 \mid v_i^2 < v_i(t_i) \le v_i^1\big\}.$$

Note that v_i^1 is a convex combination of $v_i^2 < v_i^1$ and $\mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in T_i^1 \setminus T_i^2]$. Since every $t_i \in T_i^1 \setminus T_i^2$ has $v_i(t_i) \le v_i^1$, this expectation is also below v_i^1, a contradiction.

Turning to equilibrium strategies, note that $X_i^*(s_i, \{t_i\}) = v_i(t_i)$. Also, it is easy to see that if $T_i^0 = \emptyset$, then $v_i^* = \min_{t_i \in T_i} v_i(t_i)$ and the essentially unique equilibrium has every type proving her type. This is the usual unraveling argument. So for the rest of this proof, assume $T_i^0 \neq \emptyset$.

We cannot have (s_i, T_i) and (s_i', T_i), both with positive probability in equilibrium with $X_i^*(s_i, T_i) \neq X_i^*(s_i', T_i)$ as every type strictly prefers whichever report yields the larger x. Hence we fix some s_i^* and suppose that the only (s_i, T_i) sent with positive probability in equilibrium is (s_i^*, T_i) where $X_i^*(s_i^*, T_i) \ge X_i^*(s_i, T_i)$ for all $s_i \in T_i$.

Let $\tilde{v}_i = X_i^*(s_i^*, T_i)$. Types $t_i \in T_i^0$ send report (s_i^*, T_i). Any type $t_i \notin T_i^0$ can send either (s_i^*, T_i) and obtain response $\tilde{v}_i$ or some $(s_i, \{t_i\})$ and receive response $v_i(t_i)$. Hence t_i chooses the former only if $\tilde{v}_i \ge v_i(t_i)$. Hence $\tilde{v}_i$ must be the v_i^* defined in equation (S1).

(Note that v^* is not changed if we add or remove from the set of types sending this message a type with $v_i(t_i) = v_i^*$, so it does not matter how types resolve indifference.) *Q.E.D.*

SD. PROOF OF COROLLARY 2

When $v_i^+ \leq v_i^-$, there is only one equilibrium in the auxiliary game for i, so the claim follows. When $v_i^+ > v_i^-$, however, there are (essentially) two equilibria. In one, type reports are used to separate positive types from negative types. All positive types with evidence and $v_i(t_i) > v_i^+$ prove their types, as do all negative types with evidence and $v_i(t_i) < v_i^-$. All other positive types send one type report and evidence $e_i = T_i$, while all other negative types send another type report and the same evidence. In what follows, we refer to this equilibrium as the *cheap-talk equilibrium* as it uses the "cheap talk" of type reports to help separate. In the other equilibrium, the principal's beliefs depend only on the evidence presented, so type reports are irrelevant. All positive types with evidence and $v_i(t_i) > v_i^*$ prove their types, as do all negative types with evidence and $v_i(t_i) < v_i^*$. All other types report some fixed type report and evidence $e_i = T_i$. We refer to this equilibrium as the *non-talk equilibrium*.[2]

Since there are two equilibria in the auxiliary game for i in this case, we need to determine which strategies for i are used in the equilibrium of the game which has the same outcome as the optimal mechanism. Clearly, if the principal is better off under one set of strategies than the other, then these must be the strategies used since the equilibrium corresponding to the optimal mechanism must be the best possible equilibrium for the principal.

We now show that the principal's payoff is always at least weakly larger in the cheap-talk equilibrium, completing the proof of Corollary 2.

First, we show that $v_i^+ > v_i^-$ implies $v_i^+ \geq v_i^* \geq v_i^-$ with at least one strict inequality. To see this, suppose to the contrary that $v_i^* > v_i^+ > v_i^-$. Define the following sets of types:

$$\hat{T}_i^- = \{t_i \in T_i^- \mid t_i \in T_i^0 \text{ or } v_i(t_i) \geq v_i^-\},$$

$$\hat{T}_i^+ = \{t_i \in T_i^+ \mid t_i \in T_i^0 \text{ or } v_i(t_i) \leq v_i^+\},$$

$$\hat{T}_i^{*-} = \{t_i \in T_i^- \mid t_i \in T_i^0 \text{ or } v_i(t_i) \geq v_i^*\},$$

$$\hat{T}_i^{*+} = \{t_i \in T_i^+ \mid t_i \in T_i^0 \text{ or } v_i(t_i) \leq v_i^*\}.$$

In other words, the types in $\hat{T}_i^-$ are the negative types who "pool" together in the cheap-talk equilibrium, while $\hat{T}_i^+$ is the set of positive types who pool together in this equilibrium. Similarly, $\hat{T}_i^{*-}$ and $\hat{T}_i^{*+}$ are, respectively, the set of negative and positive types who all pool together in the non-talk equilibrium. By definition,

$$v_i^- = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^-],$$

$$v_i^+ = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^+],$$

$$v_i^* = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^{*-} \cup \hat{T}_i^{*+}].$$

Hence v_i^* is a convex combination of $\mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^{*-}]$ and $\mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^{*+}]$.

[2]We refer to this as a non-talk equilibrium rather than as a babbling equilibrium since, unlike in the usual babbling equilibria in the literature, the use of evidence does enable some communication.

Since $v_i^- < v_i^*$, we see that $\hat{T}_i^{*-} \subseteq \hat{T}_i^-$. Note that if $t_i \in \hat{T}_i^-$ but $t_i \notin \hat{T}_i^{*-}$, then $v_i^- \leq v_i(t_i) < v_i^*$. Hence $v_i^- = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^-]$ is a convex combination of $\mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^{*-}]$ and the expectation of $v_i(t_i)$ for a set of types all with $v_i(t_i) \geq v_i^-$. Hence

$$v_i^* > v_i^- = \mathrm{E}_{t_i}\big[v_i(t_i) \mid t_i \in \hat{T}_i^-\big] \geq \mathrm{E}_{t_i}\big[v_i(t_i) \mid t_i \in \hat{T}_i^{*-}\big].$$

Similarly, $v_i^+ < v_i^*$ implies that $\hat{T}_i^+ \subseteq \hat{T}_i^{*+}$. Since the types in $\hat{T}_i^{*+} \setminus \hat{T}_i^+$ all satisfy $v_i^+ < v_i(t_i) \leq v_i^*$, we see that $\mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^{*+}]$ is a convex combination of $v_i^+ = \mathrm{E}_{t_i}[v_i(t_i) \mid t_i \in \hat{T}_i^+] < v_i^*$ and an expectation of $v_i(t_i)$ for a set of types with $v_i(t_i) \leq v_i^*$. Hence

$$\mathrm{E}_{t_i}\big[v_i(t_i) \mid t_i \in \hat{T}_i^{*+}\big] < v_i^*.$$

But then we have v_i^* is a convex combination of two terms which are strictly smaller than v_i^*, a contradiction. A similar argument rules out the possibility that $v_i^+ > v_i^- > v_i^*$.

Consider the game between the agents and the principal. We know that there is a robust PBE with the same outcome as in the optimal mechanism. We know i's strategy in this equilibrium must either be the one she uses in the cheap-talk equilibrium or the one she uses in the non-talk equilibrium. Fix the strategies of all agents other than i. We know these strategies are defined from the auxiliary games for these agents, independently of which strategy i uses or the principal's response to i. Thus, we can simply determine which strategy by i leads to a higher payoff for the principal.

Note that the principal's payoff for a fixed a is linear in his expectation of v_i. Hence his maximized payoff is convex in his expectation of v_i. We now show that the distribution of beliefs for the principal in the cheap-talk equilibrium is a mean-preserving spread of the distribution in the non-talk equilibrium, completing the proof. To be precise, let (σ_i^1, x_i^1) denote the cheap-talk equilibrium strategies and (σ_i^2, x_i^2) the non-talk equilibrium strategies from the auxiliary game for i. For $k = 1, 2$, define probability distributions B^k over $\mathbf{R}$ by

$$B^k(\hat{v}_i) = \rho_i\big(\{t_i \in T_i \mid X_i^k(t_i) = \hat{v}_i\}\big).$$

(Recall that $X_i^k(t_i) = x_i^k(s_i, e_i)$ for any (s_i, e_i) with $\sigma_i^k(s_i, e_i \mid t_i) > 0$ and that ρ_i is the prior over T_i.) The law of iterated expectations implies

$$\sum_{\hat{v}_i \in \mathrm{supp}(B^k)} \hat{v}_i B^k(\hat{v}_i) = \mathrm{E}_{t_i}\big[v_i(t_i)\big], \quad k = 1, 2.$$

Hence the two distributions have the same mean.

Consider any $\hat{v}_i < v_i^-$. Since $v_i^- \leq v_i^*$, for $k = 1$ or $k = 2$, we have $X_i^k(t_i) = \hat{v}_i$ if and only if there is a negative type with evidence who has $v_i(t_i) = \hat{v}_i$. Similarly, since $v_i^* \leq v_i^+$, for any $\hat{v}_i > v_i^+$, we have $X_i^k(t_i) = \hat{v}_i$ iff there is a positive type with evidence who has $v_i(t_i) = \hat{v}_i$. Hence $B^1(\hat{v}_i) = B^2(\hat{v}_i)$ for any $\hat{v}_i \notin [v_i^-, v_i^+]$.

Also, we have $B^1(\hat{v}_i) = 0$ for all $\hat{v}_i \in (v_i^-, v_i^+)$. Any type with v_i in this range either (1) is positive and chooses to induce belief v_i^+ or (2) is negative and chooses to induce belief v_i^-. Under B^2, however, many of the types generating beliefs concentrated at v_i^- or v_i^+ in the cheap-talk equilibrium instead generate beliefs in (v_i^-, v_i^+). In particular, types without evidence or types with evidence they prefer not to show induce the belief v_i^*, a positive type with evidence who has $v_i(t_i) \in (v_i^*, v_i^+)$ generates the belief $v_i(t_i)$, and similarly for negative types. Hence B^1 is a mean-preserving spread of B^2. *Q.E.D.*

SE. COUNTEREXAMPLES

In this section, we give a series of examples illustrating which assumptions are critical for which results. As explained in the text, our robustness results hinge on the independence of types across agents, our private-values assumption that agent i's utility is independent of t_{-i}, and the separability of the principal's utility function. In the following three examples, we show that if we drop any one of these assumptions, we can generate an example where there is an optimal mechanism which is deterministic and which has the same outcome as an equilibrium of the game without commitment, but where robust incentive compatibility fails.

EXAMPLE S1—Independence: Consider a simple allocation problem with two agents but where the types are correlated. Agent 1 has three types, ℓ, m, and h, while agent 2 has two types, L and H. Assume $v_i(t_i) = t_i$ and that $\ell < L < m < H < h$. The joint distribution over (t_1, t_2) is given by

	L	H
ℓ	$\frac{3}{4} - \frac{3\varepsilon}{2}$	$\frac{\varepsilon}{4}$
m	$\frac{3\varepsilon}{4}$	$\frac{\varepsilon}{4}$
h	$\frac{3\varepsilon}{4}$	$\frac{1}{4} - \frac{\varepsilon}{2}$

where $\varepsilon > 0$ but "small." Types ℓ and h have no evidence—that is, $\mathcal{E}_1(\ell) = \mathcal{E}_1(h) = \{T_1\}$. Type m has evidence $\mathcal{E}_1(m) = \{\{m\}, T_1\}$, so she can prove her type or claim to have no evidence. Both types of agent 2 have no choice but to reveal themselves. That is, $\mathcal{E}_2(L) = \{\{L\}\}$ and $\mathcal{E}_2(H) = \{\{H\}\}$.

Because types ℓ and h have the same preferences and the same evidence, there is no way for the principal to effectively screen them. More specifically, there is an optimal mechanism which gives both types the same outcome. Given this, the principal cannot do better than to pool these two types but separate them from m. For ε sufficiently small, this means that the principal will prefer giving the good to 2 if the types are (ℓ, L) or (h, L) and to 1 if the types are (ℓ, H) or (h, H). To see this, note that the principal will know 2's type but will only know that t_1 is either ℓ or h. If $t_2 = L$, then t_1 is almost certainly $\ell < L$, so the principal prefers giving the good to 2. On the other hand, if $t_2 = H$, then t_1 is almost certainly $h > H$, so the principal gives the good to 1. If $t_1 = m$, the principal will learn both agents' types and will give the good to 1 if the types are (m, L) and 2 if they are (m, H).

It is easy to see that this is achievable in an equilibrium. Take the strategies to be that the principal ignores cheap-talk statements and that type m proves her type. Given these strategies, the principal will allocate the good just as above. To see that type m will not deviate, note that if she proves her type, she gets the good with probability 3/4 as this is the conditional probability that $t_2 = L$ given $t_1 = m$. If she deviates to withholding her evidence, the principal will only give her the good if $t_2 = H$, which has conditional probability 1/4. Hence she prefers not to deviate, so this is an equilibrium.

On the other hand, it is not a robust equilibrium and the mechanism is not robustly incentive compatible. If m knew that $t_2 = H$, she would deviate to withholding her evidence.

EXAMPLE S2—Private values: There are two agents and the principal has to allocate one unit of a good to one of them. Agent 1's types are α and β, while agent 2's types are γ and δ. Types α and γ can prove themselves, while β and δ have no evidence. Agent 2's utility is 1 if he gets the good and 0 if he does not. Agent 1 is a little different and this is where we depart from our assumptions. Agent 1 wants the good if $t_2 = \delta$, but not if $t_2 = \gamma$.

Formally, $A = \{1, 2\}$ and

$$v(a, t) = u_0(a) + \sum_i u_i(a, t_1, t_2)\bar{v}_i(t_i),$$

so u_i depends on t_j. Let

$$u_2(a, t_1, t_2) = \begin{cases} 1 & \text{if } a = 2, \\ 0, & \text{otherwise.} \end{cases}$$

This satisfies simple type dependence and private values. Let

$$u_1(a, t_1, t_2) = \begin{cases} u_1(a) & \text{if } t_2 = \delta, \\ -u_1(a) & \text{if } t_2 = \gamma, \end{cases}$$

where

$$u_1(a) = \begin{cases} 1 & \text{if } a = 1, \\ 0, & \text{otherwise.} \end{cases}$$

This satisfies a natural version of simple type dependence without private values.

Note that the principal's utility from giving the good to 2 is

$$v(2, t_1, t_2) = \bar{v}_2(t_2),$$

the same as in the simple allocation problem. However, the principal's payoff to giving the good to 1 is

$$v(1, t_1, t_2) = \begin{cases} \bar{v}_1(t_1) & \text{if } t_2 = \delta, \\ -\bar{v}_1(t_1), & \text{otherwise.} \end{cases}$$

Assume that $\bar{v}_2(\gamma) > \bar{v}_1(\alpha) > \bar{v}_2(\delta) > \bar{v}_1(\beta) > 0$. So if $t_2 = \gamma$, the principal gets a negative payoff from giving the good to 1 and a positive payoff to giving the good to 2, so he prefers to give the good to 2. If $t_2 = \delta$, then he prefers giving the good to 1 if $t_1 = \alpha$ and to 2 if $t_1 = \beta$.

It is easy to see that there is an equilibrium with this outcome. Assume the principal ignores all cheap-talk messages. Types β and δ cannot prove anything. Types α and γ can prove their types, so assume their strategies are to do so. So the principal knows exactly what the types are when he sees the evidence (or lack thereof) and so can obtain his preferred allocation. Type γ's strategy is optimal since this ensures she gets the good, her favorite outcome. Type α is also choosing a best response since proving her type means she gets the good iff $t_2 = \delta$, exactly her favorite outcome. Since the optimal mechanism cannot improve on this, there is no value to commitment and no value to randomization.

However, this mechanism is not robustly incentive compatible. In particular, if $t_2 = \gamma$ but 2 does not provide proof and claims to have type δ, then type α does not want to reveal that he is type α since this will give him the good when he does not want it. (If the probability of type γ is above 1/2, this mechanism also fails to be dominant strategy incentive compatible.)

EXAMPLE S3—Additive separability: There are two agents, each with two equally likely types, denoted α and β, where the types are independent across agents. The principal's set of actions is $A = \{0, 1, 2\}$. Both agents have type-independent utility functions with

$$u_1(a) = \begin{cases} 0 & \text{if } a = 0, \\ 1 & \text{if } a = 1, \\ 3 & \text{if } a = 2, \end{cases}$$

and

$$u_2(a) = \begin{cases} 0 & \text{if } a = 0, \\ 1 & \text{if } a = 1, \\ 4 & \text{if } a = 2. \end{cases}$$

The principal's utility function takes the form $\sum_i u_i(a) v_i(t_1, t_2)$, so the only difference is that v_i depends on both types. In fact, we assume that $v_1(t_1, t_2) = 1$ for all $t \in T$, so only v_2 depends on both types. Specifically, assume that v_2 is given by the following table:

	α_2	β_2
α_1	$-1/2$	-1
β_1	-2	1

Finally, assume $\mathcal{E}_1(\alpha_1) = \{T_1\}$, $\mathcal{E}_1(\beta_1) = \{\{\beta_1\}, T_1\}$, and $\mathcal{E}_2(t_2) = \{\{t_2\}\}$. In other words, t_2 has no choice but to prove her type, α_1 cannot prove anything, and β_1 has the option to prove nothing or to prove her type.

First, consider the game. It is easy to see that it is an equilibrium for the principal to ignore any cheap talk and for β_1 to prove her type. To see this, suppose this is agent 1's strategy. Then the principal will be able to infer both agents' types correctly. The table below gives the principal's payoffs from each action as a function of the type profile:

	$a = 0$	$a = 1$	$a = 2$
(α_1, α_2)	0	1/2	-1
(α_1, β_2)	0	1/2	-1
(β_1, α_2)	0	-1	-2
(β_1, β_2)	0	2	7

Clearly, the principal chooses $a = 1$ if agent 1 reveals her type is α_1 by not presenting any evidence, chooses $a = 0$ if 1 proves her type is β_1 and 2 proves she is α_2, and chooses $a = 2$ if each i proves she is β_i, $i = 1, 2$. The only agent who could deviate is β_1. Her payoff in the equilibrium is $(1/2)(0) + (1/2)(3) = 3/2$. Her payoff if she deviates to not proving her type and pretending to be the low type is 1, so we have an equilibrium. Since the principal gets his first-best payoff, this is the same as the outcome of the best mechanism.

To see that this is not robustly incentive compatible, note that if agent 1 knew that agent 2's type was α_2, she would deviate to imitating α_1. Hence this equilibrium is not a robust PBE and the mechanism is not robustly incentive compatible.

Next, we turn to the key assumptions for our no-need-for-commitment result. As discussed in the text, one key is our assumption of simple type dependence. This assumption has two implications. First, it implies that evidence is an important part of achieving incentive compatibility—that screening via preferences is of limited value. Second, it implies that there is no value to the principal in committing to "off path" punishments that he would not follow without commitment. The following examples illustrate these points.

EXAMPLE S4—Simple type dependence 1—Screening by preferences: There is one agent with type set $T = \{\alpha, \beta\}$ where these types are equally likely. We first present a version of this example with an infinite A, contrary to our assumption that A is finite, and then discuss a finite version. Specifically, we assume the set of actions is $\{y_1, y_2\} \times \mathbf{R}_+$ and write a typical $a \in A$ as (y_k, w). The agent's utility function is

$$u(y_k, w, t) = \begin{cases} \theta(t) + w & \text{if } k = 1, \\ w, & \text{otherwise.} \end{cases}$$

As in the text, the principal's utility function takes the form $v(a, t) = u_0(a) + u(a, t)\bar{v}(t)$. Specifically, we assume $\bar{v}(t) = 1$ for all t and

$$u_0(y_k, w) = \begin{cases} -2w - c & \text{if } k = 1, \\ -2w, & \text{otherwise,} \end{cases}$$

so

$$v(y_k, w, t) = \begin{cases} \theta(t) - c - w & \text{if } k = 1, \\ -w, & \text{otherwise.} \end{cases}$$

Intuitively, the principal is deciding whether to purchase a good (y_1) or not (y_2) where the cost of the good is c and the agent's value for it is $\theta(t)$. The principal also chooses how much money w to give the agent to help induce truthful revelation.

Assume $\theta(\alpha) > \theta(\beta) + c$, $\theta(\beta) > 0$, and $c > 2\theta(\beta)$. For example, take $\theta(\alpha) = 5$, $c = 3$, and $\theta(\beta) = 1$. Finally, assume neither type of the agent has any evidence—that is, $\mathcal{E}(t) = \{T\}$ for all t.

We can write a mechanism for this problem as specifying a probability distribution over $\{y_1, y_2\}$ and a value of w as a function of the reported type t. Since w enters linearly into both the principal's and agent's utilities, there is no need to consider randomness in w. Let $p(t)$ denote the probability of $y = y_1$ given report t and let $w(t)$ denote the value of w given report t. It is not hard to show that the optimal mechanism has $p(\alpha) = 1$, $p(\beta) = 0$, $w(\alpha) = 0$, and $w(\beta) = \theta(\beta)$. In other words, the principal provides the good iff it is efficient to do so (i.e., iff $\theta(t) > c$). He bribes the β type $\theta(\beta)$ to reveal her type to save him the cost of providing her with the good.

This mechanism is deterministic. It is also trivially robustly incentive compatible since there is only one agent. However, this outcome cannot be achieved in an equilibrium of the game without commitment. To see this, note that given any belief about the agent's type, the principal will set $w = 0$. Hence both types will simply wish to maximize the probability that $y = y_1$. So the only equilibrium is where both types pool and the principal chooses his optimal y based on his prior. This is strictly worse for the principal than the optimal mechanism.

For an example with finite A and the same properties, replace $A = \{y_1, y_2\} \times \mathbf{R}_+$ with $A = \{y_1, y_2\} \times \{0, \theta(\beta)\}$. Since the optimal mechanism in the continuum case has $w \in \{0, \theta(\beta)\}$, we get the same optimal mechanism in the finite case. Also, the conclusion that the principal always sets $w = 0$ without commitment obviously still holds. Clearly, this is "non-generic," but is needed to retain the result that the optimal mechanism is deterministic. With a fine grid replacing this specification, the optimal mechanism would involve a "small" amount of randomization so that the principal could satisfy the incentive compatibility constraint with equality.

EXAMPLE S5—Simple type dependence 2—Off path punishments[3]: There is one agent with two types, α and β. The principal has three actions, a_1, a_2, and a_3, where the agent's utility function is given by

	a_1	a_2	a_3
$u(\cdot,\alpha)$	1	0	-1
$u(\cdot,\beta)$	0	1	-1

These utilities cannot satisfy simple type dependence. Note that both types strictly prefer a_1 and a_2 to a_3 but the types differ in their ranking of a_1 versus a_2, a pattern not possible under simple type dependence. As in the text, the principal's utility function takes the form $v(a,t) = u_0(a) + u(a,t)\bar{v}(t)$ where we assume $\bar{v}(t) = -1$ for all t. u_0 and the implied $v(a,t)$ are shown in the following table:

	a_1	a_2	a_3
u_0	1	1	-2
$v(\cdot,\alpha)$	0	1	-1
$v(\cdot,\beta)$	1	0	-1

The agent can prove her type or prove nothing. That is, $\mathcal{E}(t) = \{\{t\}, T\}$ for all t.

It is not hard to see that there is a mechanism which achieves the highest possible payoff for the principal for every type and which therefore must be optimal. Specifically, consider the mechanism which selects a_1 if the agent proves her type is β, a_2 if she proves her type is α, and a_3 if she proves nothing. Clearly, the agent will always prove her type to avoid a_3, so the principal's expected payoff will be 1, the highest possible. Clearly, this is an optimal mechanism and it is deterministic and (trivially) robustly incentive compatible.

On the other hand, there is no equilibrium of the game without commitment with this payoff for the principal. In the game without commitment, a_3 is a strictly dominated action for the principal, so he cannot play it at any information set. If the agent proves her type, then the principal must choose the best action for himself given the type in equilibrium. Hence, if the agent proves her type, she gets a payoff of 0. Given this, it is not hard to show that the agent never proves her type in any equilibrium. Thus, the principal gets no information in equilibrium and his payoff is $\max\{\rho(\alpha), \rho(\beta)\}$ where ρ is the principal's prior over T.

EXAMPLE S6—Principal's utility function: In this example, we show that the principal's utility function is also important for obtaining the result that commitment is not needed. As with Example S4, we use infinite A for the same reasons as discussed there. Let $A = \{y_1, y_2\} \times \mathbf{R}_+$. Assume there is only one agent with two possible types α and β where we assume the types are equally likely. Assume the agent has type-independent utility function $u(y,w) = \theta_A(y) + w$ and that the principal's utility function

[3]We thank an anonymous referee for this example.

is $v(y, w, t) = \theta_P(y, t) - \lambda w$ where

	y_1	y_2
$\theta_A(y)$	1	0
$\theta_P(y, \alpha)$	-1	1
$\theta_P(y, \beta)$	1	-1

Intuitively, think of y_1 as giving a good to the agent which the agent values at 1 and y_2 as not giving the good. The principal wants to give the good to the agent iff her type is β. The principal can use bribes of w to induce truth telling.

Assume $\mathcal{E}(\alpha) = \{\{\alpha\}, T\}$ and $\mathcal{E}(\beta) = \{T\}$, so α can prove her type but β cannot prove anything. Assume $\lambda \in (0, 1)$. Then it is not hard to show that the optimal mechanism has the principal choosing y_1 with probability 1 for type β and probability 0 for type α and setting $w = 1$ for type α and 0 for type β. In other words, the principal pays 1 to type α to reveal her type so that he can provide the good only for type β. This mechanism is deterministic and trivially is robustly incentive compatible.

In the game without commitment, though, the principal always sets $w = 0$. It is not hard to show that the principal's expected payoff in any equilibrium is 0, which is strictly worse for the principal than the optimal mechanism.

SF. ROBUST INCENTIVE COMPATIBILITY

In this section, we show the result mentioned in the conclusion that any incentive compatible mechanism is payoff equivalent to some robustly incentive compatible mechanism. That is, if P is an incentive compatible mechanism, then there is a robustly incentive compatible mechanism P^* such that the principal and every type of every agent gets the same expected utility in P and P^*.

Formally, given a mechanism P, let

$$\mathcal{U}_i(r; P) = \sum_a P(a \mid r) u_i(a)$$

and

$$\hat{\mathcal{U}}_i(r_i; P) = \mathrm{E}_{t_{-i}} \mathcal{U}_i\big(r_i, t_{-i}, M_{-i}(t_{-i}); P\big),$$

where $r \in T \times \mathcal{E}$ and $r_i \in T_i \times \mathcal{E}_i$. P and P^* are *payoff equivalent* if $\hat{\mathcal{U}}_i(t_i, M_i(t_i); P) = \hat{\mathcal{U}}_i(t_i, M_i(t_i); P^*)$ for all $t_i \in T_i$ and all i and

$$\mathrm{E}_t P\big(a \mid t, M(t)\big) = \mathrm{E}_t P^*\big(a \mid t, M(t)\big).$$

To understand the latter condition, note that if P is incentive compatible, then we can write the principal's payoff from P as

$$\begin{aligned}\mathrm{E}_t \sum_a P\big(a \mid t, M(t)\big) v(a, t) &= \mathrm{E}_t \sum_a P\big(a \mid t, M(t)\big) \Big[u_0(a) + \sum_i u_i(a) v_i(t_i) \Big] \\ &= \sum_a \mathrm{E}_t P\big(a \mid t, M(t)\big) u_0(a) + \sum_i \mathrm{E}_{t_i} \hat{\mathcal{U}}_i(t_i) v_i(t_i).\end{aligned}$$

So under the conditions above, the principal's expected payoff is the same under P and P^*.

A mechanism P is incentive compatible iff for all i,

$$\begin{aligned}\hat{\mathcal{U}}_i\big(t_i, M_i(t_i); P\big) &\geq \hat{\mathcal{U}}_i(s_i, e_i; P), \quad \forall t_i \in T_i^+, s_i \in T_i, e_i \in \mathcal{E}_i(t_i),\\ \hat{\mathcal{U}}_i\big(t_i, M_i(t_i); P\big) &\leq \hat{\mathcal{U}}_i(s_i, e_i; P), \quad \forall t_i \in T_i^-, s_i \in T_i, e_i \in \mathcal{E}_i(t_i).\end{aligned} \tag{S2}$$

This condition states that any feasible report for any positive type t_i reduces the expectation of $u_i(a)$ relative to reporting truthfully and presenting maximal evidence, while any feasible report for any negative type raises the expectation of $u_i(a)$. Hence any other feasible report for any type is weakly worse than reporting truthfully and presenting maximal evidence. Similarly, P is robustly incentive compatible iff for all i, all $s_i \in T_i$, and all $r_{-i} \in T_{-i} \times \mathcal{E}_{-i}$,

$$\begin{aligned}\mathcal{U}_i\big(t_i, M_i(t_i), r_{-i}; P\big) &\geq \mathcal{U}_i(s_i, e_i, r_{-i}; P), \quad \forall t_i \in T_i^+, e_i \in \mathcal{E}_i(t_i),\\ \mathcal{U}_i\big(t_i, M_i(t_i), r_{-i}; P\big) &\leq \mathcal{U}_i(s_i, e_i, r_{-i}; P), \quad \forall t_i \in T_i^-, e_i \in \mathcal{E}_i(t_i).\end{aligned} \tag{S3}$$

We now show that for any incentive compatible mechanism P, there is a payoff equivalent mechanism P' such that, for all $r_i, r_i' \in T_i \times \mathcal{E}_i$, $\hat{\mathcal{U}}_i(r_i; P) \geq \hat{\mathcal{U}}_i(r_i'; P)$ implies $\mathcal{U}_i(r_i, r_{-i}; P') \geq \mathcal{U}_i(r_i, r_{-i}; P')$ for all $r_{-i} \in T_{-i} \times \mathcal{E}_{-i}$. Since P is incentive compatible, it satisfies equations (S2). Hence this will show that P' satisfies equations (S3), implying that it is robustly incentive compatible.

First, to simplify the argument, note that Lemmas 3 and 4 together imply that there is a mechanism, say $\bar{P}$, which is payoff equivalent to P such that, for every $r_i \in T_i \times \mathcal{E}_i$, there is some t_i such that $\bar{P}(\cdot \mid r_i, r_{-i}) = \bar{P}(\cdot \mid t_i, M_i(t_i), r_{-i})$ for all $r_{-i} \in T_{-i} \times \mathcal{E}_{-i}$. In other words, every report is treated exactly the same way as some report of the form $(t_i, M_i(t_i))$. Hence it suffices to show that there is a mechanism P' which is payoff equivalent to P with the property that, for every $t_i, t_i' \in T_i$, $\hat{\mathcal{U}}_i(t_i, M_i(t_i); P) \geq \hat{\mathcal{U}}_i(t_i', M_i(t_i'); P)$ implies

$$\mathcal{U}_i\big(t_i, M_i(t_i), t_{-i}, M_{-i}(t_{-i}); P'\big) \geq \mathcal{U}_i\big(t_i', M_i\big(t_i'\big), t_{-i}, M_{-i}(t_{-i}); P'\big)$$

for all $t_{-i} \in T_{-i}$.

To simplify the notation, let $\mathcal{U}_i^*(t; P) = \mathcal{U}_i(t, M(t); P)$ and $\hat{\mathcal{U}}_i^*(t_i; P) = \hat{\mathcal{U}}_i(t_i, M_i(t_i); P)$. Similarly, given a mechanism P, define $P^* : T \to \Delta(A)$ by $P^*(a \mid t) = P(a \mid t, M(t))$. Using this notation, we show that if P is incentive compatible, then there exists a payoff equivalent P' with the property that $\hat{\mathcal{U}}_i^*(t_i; P) \geq \hat{\mathcal{U}}_i^*(t_i'; P)$ implies $\mathcal{U}_i^*(t_i, t_{-i}; P') \geq \mathcal{U}_i^*(t_i', t_{-i}; P')$ for all $t_{-i} \in T_{-i}$.

This proof is essentially the same as the proof of Lemma 1 in Gershkov et al. (2013). We present it here for completeness.

Given P, let $\bar{P}$ solve the problem of minimizing the variance of agents' utilities subject to being payoff equivalent to P. That is, $\bar{P}$ minimizes

$$\mathrm{E}_t \sum_i \big[\mathcal{U}_i^*(t; \bar{P})\big]^2$$

subject to $\bar{P}$ being payoff equivalent to P. (Existence of a minimizer follows from continuity of the objective function and compactness of the set of mechanisms satisfying the constraint.) We now show that $\hat{\mathcal{U}}_i^*(t_i; \bar{P}) \geq \hat{\mathcal{U}}_i^*(t_i'; \bar{P})$ implies $\mathcal{U}_i^*(t_i, t_{-i}; \bar{P}) \geq \mathcal{U}_i^*(t_i', t_{-i}; \bar{P})$.

To see this, suppose not. Then there exist i, $\bar{t}_i, \bar{t}_i' \in T_i$ and $\bar{t}_{-i} \in T_{-i}$ with $\hat{\mathcal{U}}_i^*(\bar{t}_i; \bar{P}) \leq \hat{\mathcal{U}}_i^*(\bar{t}_i'; \bar{P})$ but $\mathcal{U}_i^*(\bar{t}_i, \bar{t}_{-i}; \bar{P}) > \mathcal{U}_i^*(\bar{t}_i', \bar{t}_{-i}; \bar{P})$. Clearly, this implies that there must be some $\bar{t}_{-i}' \in T_{-i}$ such that $\mathcal{U}_i^*(\bar{t}_i, \bar{t}_{-i}'; \bar{P}) < \mathcal{U}_i^*(\bar{t}_i', \bar{t}_{-i}'; \bar{P})$.

Without loss of generality, assume the prior probability of the four type profiles $\bar{t}$, $\bar{t}'$, $(\bar{t}_i, \bar{t}_{-i}')$, and $(\bar{t}_i', \bar{t}_{-i})$ are the same.[4] Let $\bar{T}$ denote this set of type profiles.

Fix a small $\varepsilon > 0$ and let

$$\lambda = \frac{\varepsilon}{\mathcal{U}_i^*(\bar{t}; \bar{P}) - \mathcal{U}_i^*(\bar{t}_i', \bar{t}_{-i}; \bar{P})},$$
$$\lambda' = \frac{\varepsilon}{\mathcal{U}_i^*(\bar{t}'; \bar{P}) - \mathcal{U}_i^*(\bar{t}_i, \bar{t}_{-i}'; \bar{P})}.$$

Both denominators are strictly positive by hypothesis, so by choosing ε sufficiently small, we guarantee $\lambda, \lambda' \in (0, 1)$.

Consider a new mechanism, $\hat{P}$, where

$$\begin{aligned}
\hat{P}^*(\cdot \mid \bar{t}) &= (1-\lambda)\bar{P}^*(\cdot \mid \bar{t}) + \lambda \bar{P}^*(\cdot \mid \bar{t}_i', \bar{t}_{-i}), \\
\hat{P}^*(\cdot \mid \bar{t}_i', \bar{t}_{-i}) &= (1-\lambda)\bar{P}^*(\cdot \mid \bar{t}_i', \bar{t}_{-i}) + \lambda \bar{P}^*(\cdot \mid \bar{t}), \\
\hat{P}^*(\cdot \mid \bar{t}_i, \bar{t}_{-i}') &= (1-\lambda')\bar{P}^*(\cdot \mid \bar{t}_i, \bar{t}_{-i}') + \lambda' \bar{P}^*(\cdot \mid \bar{t}'), \\
\hat{P}^*(\cdot \mid \bar{t}') &= (1-\lambda')\bar{P}^*(\cdot \mid \bar{t}') + \lambda' \bar{P}^*(\cdot \mid \bar{t}_i, \bar{t}_{-i}'),
\end{aligned}$$

and $\hat{P}(\cdot \mid r) = \bar{P}(\cdot \mid r)$ for all other $r \in T \times \mathcal{E}$. To see that $\hat{P}$ is payoff equivalent to $\bar{P}$ and therefore to P, note that

$$\mathrm{E}_t \hat{P}^*(a \mid t) = \Pr[\bar{T}]\mathrm{E}_t\big[\hat{P}^*(a \mid t) \mid t \in \bar{T}\big] + \big(1 - \Pr[\bar{T}]\big)\mathrm{E}_t\big[\hat{P}^*(a \mid t) \mid t \notin \bar{T}\big].$$

Since $\hat{P}$ is the same as $\bar{P}$ for $t \notin \bar{T}$, we have $\mathrm{E}_t[\hat{P}^*(a \mid t) \mid t \notin \bar{T}] = \mathrm{E}_t[\bar{P}^*(a \mid t) \mid t \notin \bar{T}]$. Since the type profiles in $\bar{T}$ are equally likely,

$$\begin{aligned}
\mathrm{E}_t\big[\hat{P}^*(a \mid t) \mid t \in \bar{T}\big] &= \frac{1}{4}\big[\hat{P}^*(a \mid \bar{t}) + \hat{P}^*(a \mid \bar{t}_i', \bar{t}_{-i}) + \hat{P}^*(a \mid \bar{t}_i, \bar{t}_{-i}') + \hat{P}^*(a \mid \bar{t}')\big] \\
&= \frac{1}{4}\big[\bar{P}^*(a \mid \bar{t}) + \bar{P}^*(a \mid \bar{t}_i', \bar{t}_{-i}) + \bar{P}^*(a \mid \bar{t}_i, \bar{t}_{-i}') + \bar{P}^*(a \mid \bar{t}')\big],
\end{aligned}$$

so $\mathrm{E}_t \hat{P}^*(a \mid t) = \mathrm{E}_t \bar{P}^*(a \mid t)$. Also, clearly, for $t_i \notin \{\bar{t}_i, \bar{t}_i'\}$, we have $\hat{\mathcal{U}}_i^*(t_i; \hat{P}) = \hat{\mathcal{U}}_i^*(t_i; \bar{P})$ since the mechanisms treat such t_i the same way. For other types, we have

$$\begin{aligned}
\hat{\mathcal{U}}^*(\bar{t}_i; \hat{P}) &= \mathrm{E}_{t_{-i}} \mathcal{U}_i^*(\bar{t}_i, t_{-i}; \hat{P}) \\
&= \Pr[\bar{T}_{-i}]\mathrm{E}_{t_{-i}}\big[\mathcal{U}_i^*(\bar{t}_i, t_{-i}; \hat{P}) \mid t_{-i} \in \bar{T}_{-i}\big] \\
&\quad + \big(1 - \Pr[\bar{T}_{-i}]\big)\mathrm{E}_{t_{-i}}\big[\mathcal{U}_i^*(\bar{t}_i, t_{-i}; \hat{P}) \mid t_{-i} \notin \bar{T}_{-i}\big].
\end{aligned}$$

[4]If this is not true, we can split each type into two equivalent types where one of the new types has some fixed probability. Then we can apply the argument to these new type profiles.

Again, $\hat{\mathcal{U}}_i^*(\bar{t}_i, t_{-i}; \hat{P}) = \hat{\mathcal{U}}_i^*(\bar{t}_i, t_{-i}; \bar{P})$ if $t_{-i} \notin \bar{T}_{-i}$. Also,

$$\mathrm{E}_{t_{-i}}\big[\hat{\mathcal{U}}_i^*(\bar{t}_i, t_{-i}; \hat{P}) \mid t_{-i} \in \bar{T}_{-i}\big] = \frac{1}{2}\big[\hat{\mathcal{U}}_i^*(\bar{t}_i, \bar{t}_{-i}; \hat{P}) + \hat{\mathcal{U}}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \hat{P}\big)\big].$$

Substituting from the definition, this is

$$\frac{1}{2}\big[\mathcal{U}_i^*(\bar{t}; \bar{P}) + \lambda\big(\mathcal{U}_i^*\big(\bar{t}'_i, \bar{t}_{-i}; \bar{P}\big) - \mathcal{U}_i^*(\bar{t}; \bar{P})\big) + \mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \bar{P}\big) + \lambda'\big(\mathcal{U}_i^*\big(\bar{t}'; \bar{P}\big) - \mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \bar{P}\big)\big)\big].$$

Using the definition of λ and λ', this is

$$\frac{1}{2}\big[\mathcal{U}_i^*(\bar{t}; \bar{P}) + \mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \bar{P}\big) - \varepsilon + \varepsilon\big] = \frac{1}{2}\big[\mathcal{U}_i^*(\bar{t}; \bar{P}) + \mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \bar{P}\big)\big],$$

which is $\mathrm{E}_{t_{-i}}[\mathcal{U}_i^*(\bar{t}_i, t_{-i}; \bar{P}) \mid t_{-i} \in \bar{T}_{-i}]$. Hence $\hat{\mathcal{U}}_i^*(\bar{t}_i; \hat{P}) = \hat{\mathcal{U}}_i^*(\bar{t}_i; \bar{P})$. The calculation for $\bar{t}'_i$ is analogous.

So $\hat{P}$ is payoff equivalent to $\bar{P}$. We now show that $\hat{P}$ has strictly lower variance, a contradiction that establishes the result. Note that

$$\begin{aligned}
\mathrm{E}_t \sum_i \big[\mathcal{U}_i^*(t; \hat{P})\big]^2 &= \Pr[\bar{T}]\mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \hat{P})\big]^2 \mid t \in \bar{T}\Big) \\
&\quad + \big(1 - \Pr[\bar{T}]\big)\mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \hat{P})\big]^2 \mid t \notin \bar{T}\Big) \\
&= \Pr[\bar{T}]\mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \hat{P})\big]^2 \mid t \in \bar{T}\Big) \\
&\quad + \big(1 - \Pr[\bar{T}]\big)\mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \bar{P})\big]^2 \mid t \notin \bar{T}\Big),
\end{aligned}$$

so it suffices to show that

$$\mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \hat{P})\big]^2 \mid t \in \bar{T}\Big) < \mathrm{E}_t\Big(\sum_i \big[\mathcal{U}_i^*(t; \bar{P})\big]^2 \mid t \in \bar{T}\Big). \tag{S4}$$

It is easy to see that the left-hand side is 1/4 times

$$\begin{aligned}
&\big[(1-\lambda)^2 + \lambda^2\big]\sum_i \big[\mathcal{U}_i^*(\bar{t}; \hat{P})\big]^2 + \big[(1-\lambda)^2 + \lambda^2\big]\sum_i \big[\mathcal{U}_i^*\big(\bar{t}'_i, \bar{t}_{-i}; \hat{P}\big)\big]^2 \\
&\quad + \big[\big(1-\lambda'\big)^2 + \big(\lambda'\big)^2\big]\sum_i \big[\mathcal{U}_i^*\big(\bar{t}'; \hat{P}\big)\big]^2 + \big[\big(1-\lambda'\big)^2 + \big(\lambda'\big)^2\big]\sum_i \big[\mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \hat{P}\big)\big]^2 \\
&\quad + 4\lambda(1-\lambda)\Big(\sum_i \mathcal{U}_i^*(\bar{t}; \hat{P})\Big)\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}'_i, \bar{t}_{-i}; \hat{P}\big)\Big) \\
&\quad + 4\lambda'\big(1-\lambda'\big)\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}'; \hat{P}\big)\Big)\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}_i, \bar{t}'_{-i}; \hat{P}\big)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t\in\bar{T}} \sum_i \big[\mathcal{U}_i^*(t;\bar{P})\big]^2 \\
&\quad - 2\lambda(1-\lambda)\bigg[\sum_i \big[\mathcal{U}_i^*(\bar{t};\hat{P})\big]^2 \\
&\quad - 2\Big(\sum_i \mathcal{U}_i^*(\bar{t};\hat{P})\Big)\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}_i',\bar{t}_{-i};\hat{P}\big)\Big) + \sum_i \big[\mathcal{U}_i^*\big(\bar{t}_i',\bar{t}_{-i};\hat{P}\big)\big]^2\bigg] \\
&\quad - 2\lambda'\big(1-\lambda'\big)\bigg[\sum_i \big[\mathcal{U}_i^*\big(\bar{t}';\hat{P}\big)\big]^2 \\
&\quad - 2\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}';\hat{P}\big)\Big)\Big(\sum_i \mathcal{U}_i^*\big(\bar{t}_i,\bar{t}_{-i}';\hat{P}\big)\Big) + \sum_i \big[\mathcal{U}_i^*\big(\bar{t}_i,\bar{t}_{-i}';\hat{P}\big)\big]^2\bigg] \\
&= \sum_{t\in\bar{T}} \sum_i \big[\mathcal{U}_i^*(t;\bar{P})\big]^2 - 2\lambda(1-\lambda)\sum_i \big[\mathcal{U}_i^*(\bar{t};\hat{P}) - \mathcal{U}_i^*\big(\bar{t}_i',\bar{t}_{-i};\hat{P}\big)\big]^2 \\
&\quad - 2\lambda'\big(1-\lambda'\big)\sum_i \big[\mathcal{U}_i^*\big(\bar{t}';\hat{P}\big) - \mathcal{U}_i^*\big(\bar{t}_i,\bar{t}_{-i}';\hat{P}\big)\big]^2,
\end{aligned}
$$

which implies equation (S4).

REFERENCES

FUDENBERG, D., AND J. TIROLE (1991): “Perfect Bayesian Equilibrium and Sequential Equilibrium,” *Journal of Economic Theory*, 53, 236–260. [2]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 9 January, 2017; final version accepted 31 October, 2018; available online 2 November, 2018.