

# CORRELATED PERFECT EQUILIBRIUM

WANYING HUANG, J. JUDE KLINE, AND PRISCILLA MAN

**ABSTRACT.** We propose a refinement of correlated equilibrium based on mediator errors, called *correlated perfect equilibrium* (CPE). In finite games, the set of CPE is nonempty and forms a finite union of convex sets. Like perfect equilibrium, a CPE never assigns positive probability to any weakly dominated strategy. We provide a dual representation of CPE and demonstrate how it differs from two existing refinements of correlated equilibrium—acceptable correlated equilibrium (Myerson, 1986) and perfect direct correlated equilibrium (Dhillon–Mertens, 1996)—through examples.

**Keywords:** Correlated Equilibrium, Perfection, Refinement, Normal Form Games

**JEL classification codes:** C72

---

*Date:* October 9, 2025.

A version of part of this paper was submitted to the University of Queensland as Huang’s thesis in partial fulfillment of the Honours degree requirement. The authors thank seminar participants at University of Bristol, Virginia Tech, the 2020 Australasian Economic Theory Workshop, and FUR 2024. Man acknowledges the support from the Australia Research Council Discovery Early Career Researcher Award DE180101452.

Department of Economics, Monash University. E-mail: [kate.huang@monash.edu](mailto:kate.huang@monash.edu).

School of Economics, University of Queensland. E-mail: [j.kline@uq.edu.au](mailto:j.kline@uq.edu.au).

School of Economics, University of Queensland. E-mail: [t.man@uq.edu.au](mailto:t.man@uq.edu.au).

## 1. INTRODUCTION

A central solution concept in game theory is correlated equilibrium (Aumann, 1974).<sup>1</sup> It is often described as a coordination plan in which a mediator draws a joint strategy profile from a distribution and privately recommends to each player her component. While each player only observes her own recommendation, the distribution from which each recommendation is drawn is common knowledge. This interpretation assumes that the mediator never errs—that is, recommendations are always transmitted exactly as drawn from the distribution. In practice, coordination devices—such as matching algorithms or communication channels—are rarely error-free. This raises a natural question: how should correlated equilibrium be defined when the device itself is imperfect?

In this paper, we propose a new solution concept, the *correlated perfect equilibrium* (CPE), which allows for the possibility that the mediator makes mistakes. Formally, a CPE is a distribution over strategy profiles that arises as the limit of a sequence of completely mixed distributions; in such a sequence, any strategy recommended with positive probability in the limit must remain a best response throughout. Since every strategy receives positive probability along the sequence, off-equilibrium recommendations can be interpreted as mediator errors. Correlated perfection therefore requires that, conditional on receiving such a recommendation, players have no unilateral incentive to deviate from the mediator’s advice. By attributing mistakes to the mediator, CPE also ensures that all players share the same beliefs about the error process, thus guaranteeing consistency of off-equilibrium beliefs.

We obtain the following results. As a refinement of correlated equilibrium, we establish the existence of CPE in finite normal-form games. In line with Selten (1975)’s perfect equilibrium, a CPE never assigns positive probability to weakly dominated strategies; in fact, every perfect equilibrium is itself a CPE. Our main result (Theorem 1) provides a dual characterization of the set of CPEs. It then follows that the set of CPEs can be expressed as a finite union of convex polyhedra. The dual characterization also yields a tractable iterative method for identifying these equilibria. Moreover, it offers an alternative interpretation of CPE: it is a correlated equilibrium that is robust not only to players’ unilateral deviations but also to any *collectively profitable* deviation plan—namely, any joint deviation where the aggregate gains of some players strictly outweigh the losses of others.

The primary advantages of CPE stem from its conceptual simplicity. By attributing errors to the mediator—or by viewing the mediator’s private recommendations as being transmitted through a noisy channel—mistakes originate from a single, common

---

<sup>1</sup>It can be understood either as an expression of Bayesian rationality with a common prior (Aumann, 1987), or as a condition that rules out joint “Dutch books” and thus precludes arbitrage opportunities for an outside observer (Nau and McCardle, 1990). As a generalization of Nash equilibrium, the correlated equilibrium is also much easier to compute; see, e.g., Papadimitriou and Roughgarden (2008); Daskalakis et al. (2009).

source. This perspective yields two key features of CPE that distinguishes it from existing notions. First, it accommodates correlated errors without requiring any additional assumptions about their relative likelihood, while ensuring consistency of off-equilibrium beliefs. Second, it is invariant to the duplication of payoff-identical strategies, so the refinement depends only on the underlying strategic environment rather than its particular representation. In Section 4, we illustrate these features through two examples. These examples also highlight how CPE differs from other refinements of correlated equilibrium, which we discuss next.

**Related Literature.** Existing refinements of correlated equilibrium have primarily focused on perturbations of players’ actions.<sup>2</sup> The first such refinement is Myerson (1986)’s *acceptable correlated equilibrium* (ACE), which adapts Selten’s idea of perfection to allow for correlated player mistakes. However, because players were not informed of their own trembles, in ACE, agents could form beliefs about others’ mistakes that were inconsistent with the information available to them. This concern is raised by Dhillon and Mertens (1996). In response, they propose the notion of *perfect correlated equilibrium* (PCE), which is defined as a perfect equilibrium in the extended communication game.<sup>3</sup> This framework imposes independence across players’ mistakes, which in turn resolves the issue of inconsistent beliefs; moreover, Dhillon and Mertens show that every PCE is also an ACE.

A drawback of PCE, however, is that it fails to satisfy the revelation principle, complicating its use in applications.<sup>4</sup> To address this, more recently, Luo, Qiao, and Sun (2022) introduce the *correlated equilibrium with message-dependent trembles* ( $\text{CE}^{\text{MDT}}$ ), which allows players to hold subjective, message-dependent beliefs about others’ mistakes. Unlike PCE,  $\text{CE}^{\text{MDT}}$  satisfies the revelation principle: every  $\text{CE}^{\text{MDT}}$  distribution can be implemented by the direct mechanism. In this setting—which is also our focus—Luo et al. (2022) provides a general framework that both nests PCE as a special case and coincides with a weak version of ACE.<sup>5</sup> As we will see in Section 4, our notion of CPE is distinct from both, as there are no set-inclusion relationships between CPE and either PCE or ACE.

Conceptually, our notion departs from the player-tremble tradition by attributing mistakes to the mediator. There is a single common source of error, so all players share the same belief distribution over these errors. This avoids the belief-inconsistency problem that arises under ACE, while still allowing correlation through an objective error process. By contrast,  $\text{CE}^{\text{MDT}}$  adopts a subjective perspective, permitting heterogeneous

<sup>2</sup>There is another strand of literature that generalizes correlated equilibrium, e.g., see Moulin and Vial (1978); Grant and Stauber (2022).

<sup>3</sup>While the message space in an extended communication game need not coincide with the players’ strategy space, we focus on the direct mechanism where they do.

<sup>4</sup>Formally, this means that there is a PCE that cannot be induced by a PCE via a direct mechanism.

<sup>5</sup>See their Proposition 3.

beliefs about tremble probabilities across players. While our notion coincides with  $\text{CE}^{\text{MDT}}$  when all players happen to share the same subjective belief, our focus is different. The primary goal of [Luo et al. \(2022\)](#) is to restore the revelation principle, whereas our work provides a distinct rationale for the common-belief structure—deriving it from mediator error—and a tractable characterization of the resulting equilibria.<sup>6</sup>

## 2. PRELIMINARIES

Consider a game  $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ , where  $N$  is a finite set of players,  $S_i$  is a finite set of pure strategies of player  $i$ ,  $S = \prod_{i \in N} S_i$  is the set of pure strategy profiles, and the function  $u_i : S \rightarrow \mathbb{R}$  gives player  $i$ 's payoff at each strategy profile  $s \in S$ .

A correlated strategy  $\rho$  is a probability distribution over  $S$  and we say that it is *completely mixed* if it assigns positive probability to each element of  $S$ . Following [Aumann \(1974, 1987\)](#), a correlated strategy  $\rho$  is a *correlated equilibrium* if for each player  $i \in N$  and for every strategy  $s_i \in S_i$ , we have

$$\sum_{s_{-i} \in S_{-i}} \rho(s_i, s_{-i}) \cdot u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \rho(s_i, s_{-i}) \cdot u_i(s'_i, s_{-i}) \quad \text{for all } s'_i \in S_i. \quad (1)$$

A correlated equilibrium  $\rho$  is usually interpreted as follows: a mediator draws a strategy profile  $s$  according to  $\rho$  and privately recommends to each player  $i$  her component  $s_i$ . Condition (1) requires that, given the distribution  $\rho$  and assuming all opponents follow their recommendations, each player finds it optimal to follow her own recommendation.

It follows directly from (1) that the set of correlated equilibria is a compact, convex subset of  $\Delta S$  that contains all Nash equilibria. The existence of correlated equilibria follows from that of a Nash equilibrium ([Nash, 1950](#)).<sup>7</sup>

**Correlated Perfect Equilibrium.** We now introduce our notion of perfection for correlated equilibrium. In analogy to perfect equilibrium ([Selten, 1975](#)), where players may independently make arbitrarily small mistakes, we instead allow the mediator to err when issuing private recommendations. Since these mistakes come from a common source—namely, the mediator—they can be correlated across players.

Formally, for a correlated strategy  $\rho$ , let  $\rho_i(s_i)$  denote the unconditional probability that the mediator recommends  $s_i$ :

$$\rho_i(s_i) := \sum_{s_{-i} \in S_{-i}} \rho(s_i, s_{-i}). \quad (2)$$

Thus,  $\rho_i(s_i)$  is the ex-ante probability that player  $i$  receives recommendation  $s_i$ . We write  $S_i^\rho = \{s_i \in S_i : \rho_i(s_i) > 0\}$  for the set of  $i$ 's strategies that lie in the support of  $\rho$ , and  $S^\rho = \prod_{i \in N} S_i^\rho$  for the product of these sets across all players. Note that  $S^\rho$  is the smallest product set of strategy profiles containing the support of  $\rho$ .

<sup>6</sup>As discussed in Section 5.1 of [Luo et al. \(2022\)](#), the revelation principle no longer holds in our setting.

<sup>7</sup>For a direct proof of the existence of correlated equilibria, see [Hart and Schmeidler \(1989\)](#).

We say that a correlated strategy  $\rho \in \Delta S$  is a *correlated perfect equilibrium* if there exists a sequence of completely mixed correlated strategies  $(\rho^k)$  converging to  $\rho$  such that, for each player  $i \in N$  and every strategy  $s_i \in S_i^\rho$ , we have

$$\sum_{s_{-i} \in S_{-i}} \rho^k(s_i, s_{-i}) \cdot u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \rho^k(s_i, s_{-i}) \cdot u_i(s'_i, s_{-i}) \quad \text{for all } s'_i \in S_i, k \in \mathbb{N}. \quad (3)$$

Clearly, every correlated perfect equilibrium is a correlated equilibrium, so our notion is a refinement. Intuitively, a correlated perfect equilibrium is a correlated equilibrium that is robust to arbitrarily small and possibly correlated errors in the mediator's recommendations. That is, for each player  $i$ , any strategy  $s_i$  that is recommended with positive probability in the limit—i.e.,  $s_i \in S_i^\rho$ —must be a best response under such perturbations.

Notice that any correlated equilibrium that is completely mixed is also correlated perfect. Moreover, since correlated equilibrium generalizes Nash equilibrium, our notion naturally extends [Selten \(1975\)](#)'s perfect equilibrium, as we show below.

**Lemma 1.** Every perfect equilibrium is a correlated perfect equilibrium.

This lemma holds since one can construct a sequence of mediator mistakes that mimic the independent mistakes made by the players. Therefore, as the set of perfect equilibria is non-empty ([Selten, 1975](#)), the set of correlated perfect equilibria is non-empty as well. Beyond existence, correlated perfection also inherits another desirable property of perfect equilibrium: it rules out weakly dominated strategies.

**Lemma 2.** In any correlated perfect equilibrium, no weakly dominated strategy is assigned positive probability.

Recall that, just as with Nash equilibrium, a correlated equilibrium may assign positive probability to weakly dominated strategies. Thus, [Lemma 2](#) implies that the set of correlated perfect equilibria is contained within the set of correlated equilibria that assign zero probability to all weakly dominated strategies.

### 3. A DUAL REPRESENTATION OF CORRELATED PERFECT EQUILIBRIUM

In this section, we provide a characterization of the set of correlated perfect equilibria. By definition, verifying that a correlated equilibrium is correlated perfect requires finding only a single supporting sequence that satisfies condition (3). However, ruling out correlated perfection is more demanding, as one must show that no such sequence exists. To address this challenge, we introduce a dual representation. This characterization provides not only an alternative interpretation of correlated perfection but also a systematic method for identifying the entire set of these equilibria.

To formalize this, we first introduce the notion of dual vectors. For each player  $i$ , consider a vector

$$\alpha_i = (\alpha_i(s'_i \mid s_i))_{s_i, s'_i \in S_i} \in \mathbb{R}_+^{|S_i \times S_i|}$$

such that for each  $s_i \in S_i$ ,

$$\sum_{s'_i \in S_i} \alpha_i(s'_i | s_i) = 1.$$

The vector  $\alpha_i$  can be interpreted as a randomized deviation plan for player  $i$ : if the mediator recommends strategy  $s_i$ , player  $i$  deviates to  $s'_i$  with probability  $\alpha_i(s'_i | s_i)$ . For any strategy profile  $s$ , the expected gain to player  $i$  from following deviation plan  $\alpha_i$  is

$$D_i(s, \alpha_i) := \sum_{s'_i \in S_i} \alpha_i(s'_i | s_i) (u_i(s'_i, s_{-i}) - u_i(s_i, s_{-i})).$$

We call a profile of deviation plans  $\alpha = (\alpha_i)_{i \in N}$  a *dual vector* if for all  $s \in S$ , the sum of all players' expected gains from their respective deviations is nonnegative:

$$\sum_{i \in N} D_i(s, \alpha_i) \geq 0. \quad (4)$$

In other words, a dual vector specifies deviation plans that together yield a (weak) aggregate utility gain.

For a given correlated equilibrium  $\rho$ , recall that  $S^\rho$  denotes the smallest product set containing the support of  $\rho$ . We say that a dual vector  $\alpha = (\alpha_i)_{i \in N}$  is  $S^\rho$ -restricted if, for each player  $i \in N$ , the deviation plan  $\alpha_i$  always prescribes the agent to follow the mediator's recommendation whenever the recommended strategy  $s_i$  is not in  $S_i^\rho$ :  $\alpha_i(\cdot | s_i) = \delta_{s_i}$  for all  $s_i \notin S_i^\rho$ .<sup>8</sup> We focus on restricted dual vectors because strategies outside  $S_i^\rho$  are never recommended in the limit and thus irrelevant for testing whether  $\rho$  is correlated perfect.

Now, we are ready to state our main result.

**Theorem 1.** A correlated equilibrium  $\rho$  is correlated perfect if and only if for every  $S^\rho$ -restricted dual vector  $\alpha$  and every  $s \in S$ , (4) holds with equality.

Theorem 1 shows that whether a correlated equilibrium  $\rho$  is correlated perfect depends only on its minimal product support  $S^\rho$ . A direct implication is that if two correlated equilibria  $\rho$  and  $\rho'$  share the same product support ( $S^\rho = S^{\rho'}$ ), then either both are correlated perfect or neither is. Moreover, since the set of correlated equilibria supported on any fixed product set of strategies forms a convex polyhedron, the set of correlated perfect equilibria can be expressed as a finite union of convex polyhedra. However, a union of convex sets is not necessarily convex: as we will see in Example 1, a mixture of two correlated perfect equilibria could fail to be correlated perfect.

Beyond this structural property, the characterization has a clear intuitive consequence: correlated perfection requires that no restricted deviation plan yields a strictly positive aggregate utility gain. Thus, our notion immediately rules out any (restricted) plan that constitutes a strict Pareto improvement—that is, one that strictly benefits at least one player without harming any other. The condition is, in fact, more demanding: by

<sup>8</sup>We use  $\delta_{s_i}$  to denote the degenerate distribution that puts all probability mass on  $s_i$ .

requiring the sum of players' gains to be exactly zero, it also excludes plans where some players' gains outweigh others' losses. Thus, one can view a correlated perfect equilibrium as being robust not only to unilateral deviations but also to any collectively profitable deviation plan.

Finally, observe that an  $S^\rho$ -restricted dual vector  $\alpha$  is also an  $S'$ -restricted dual vector whenever  $S^\rho \subseteq S'$ . As a consequence, if the equality in Theorem 1 holds for  $S'$ , it also holds for every subset of  $S'$ . Thus, if a correlated equilibrium  $\rho$  is not correlated perfect, then by Theorem 1 any correlated equilibrium  $\rho'$  with  $S^{\rho'} \supseteq S^\rho$  is also not correlated perfect. This observation yields an iterative procedure to compute the set of correlated perfect equilibria: begin with the entire strategy space  $S$  and check whether condition (4) holds with equality for all (unrestricted) dual vectors and all  $s \in S$ . If so, then every correlated equilibrium is correlated perfect. Otherwise, restrict attention to a smaller product set  $S'$  and repeat the previous procedure. Once a correlated perfect equilibrium  $\rho$  with  $S^\rho = S'$  is identified, every correlated equilibrium supported on  $S'$  is also correlated perfect. The process terminates once every remaining product set is contained in one that is already associated with a correlated perfect equilibrium.

#### 4. EXAMPLES

In this section, we present two examples to illustrate how to identify a CPE, using both its definition and the characterization in Theorem 1. These examples also clarify the relationship between CPE and two refinements of correlated equilibrium: acceptable correlated equilibrium (ACE) (Myerson, 1986) and perfect direct correlated equilibrium (PDCE) (Dhillon and Mertens, 1996).<sup>9</sup>

Unlike our notion of perfection, in both ACE and PDCE, it is the players—not the mediator—who make mistakes. More specifically, in ACE, mistakes may be correlated across any subset of players, whereas in PDCE they must be independent. Thus, every PDCE is an ACE (see Proposition 3, Dhillon and Mertens (1996)). Nevertheless, as we show below, there is no set-inclusion relationship between our notion and these refinements. Example 1 shows a CPE that is not an ACE (and thus not a PDCE), whereas Example 2 shows a PDCE (and thus an ACE) that is not a CPE.

**Example 1** (A CPE that is not an ACE). Consider a three-person game in Figure 1a where  $S_1 = \{x_1, y_1, z_1\}$ ,  $S_2 = \{x_2, y_2, z_2\}$  and  $S_3 = \{x_3, y_3\}$ . As shown by Myerson (1986, p. 143), the *unique* acceptable correlated equilibrium in this game is the degenerate distribution that assigns probability one to strategy profile  $(x_1, x_2, x_3)$ .

In contrast, we show that the degenerate distribution  $\delta_{(y_1, y_2, y_3)}$ , which assigns probability one to  $(y_1, y_2, y_3)$ , is a correlated perfect equilibrium. To see this, we construct a completely mixed sequence of correlated strategies that converges to  $\delta_{(y_1, y_2, y_3)}$ . This

<sup>9</sup>Recall that Dhillon and Mertens (1996) allow for general correlation devices where the message space need not coincide with their strategy space. To make our notion comparable with theirs, we focus on the perfect direct correlated equilibrium which is the PCE induced via a direct mechanism.



	$x_2$	$y_2$	$z_2$		$x_2$	$y_2$	$z_2$
$x_1$	1, 1, 2	2, 0, 0	2, 0, 0	$x_1$	3, 3, 1	3, 3, 1	3, 3, 1
$y_1$	0, 2, 0	3, 0, 0	0, 3, 0	$y_1$	3, 3, 1	3, 3, 1	3, 3, 1
$z_1$	0, 2, 0	0, 3, 0	3, 0, 0	$z_1$	3, 3, 1	3, 3, 1	3, 3, 1
	$x_3$				$y_3$		

(a) Myerson's (1986) three-person game

	$x_2$	$y_2$	$z_2$		$x_2$	$y_2$	$z_2$
$x_1$	$\varepsilon$	$\varepsilon$	$\varepsilon$	$x_1$	$\varepsilon$	$\varepsilon$	$\varepsilon$
$y_1$	$\varepsilon$	$3\varepsilon$	$\varepsilon$	$y_1$	$\varepsilon$	1	$\varepsilon$
$z_1$	$\varepsilon$	$7\varepsilon$	$\varepsilon$	$z_1$	$\varepsilon$	$\varepsilon$	$\varepsilon$
	$x_3$				$y_3$		

(b) A supporting sequence for  $\delta_{(y_1, y_2, y_3)}$  to be correlated perfect.

FIGURE 1. A correlated perfect equilibrium that is not an ACE.

sequence is shown in Figure 1b.<sup>10</sup> For each  $\varepsilon > 0$  in this sequence, it is straightforward to verify that each recommended strategy is a best response for every player. For example, if player 1 is recommended  $y_1$ , her conditional expected payoff from choosing  $x_1, y_1$  and  $z_1$  are  $15\varepsilon + 3$ ,  $15\varepsilon + 3$  and  $9\varepsilon + 3$ , respectively; thus following the recommendation  $y_1$  is indeed optimal. An analogous calculation shows that  $y_2$  is optimal for player 2. For player 3,  $y_3$  yields a strictly higher expected payoff than  $x_3$  and is therefore optimal. By definition,  $\delta_{(y_1, y_2, y_3)}$  is a correlated perfect equilibrium.

**Why CPE  $\not\subseteq$  ACE.** In Example 1, we identified a CPE that is not an ACE. More specifically, notice that the distribution in Figure 1b, which supports  $\delta_{(y_1, y_2, y_3)}$  as a CPE, cannot support it as an ACE. This is because, although ACE permits correlated mistakes across players, it imposes a strict ordering of mistake probabilities: deviations by a larger group of players must be infinitesimally less likely than those by a smaller group. The distribution in Figure 1b violates this requirement as it assigns the same order of probability to profiles with different numbers of deviators—for example,  $(y_1, y_2, x_3)$  involves one deviator, namely player 3, whereas the profile  $(z_1, y_2, x_3)$  involves two deviators, i.e., players 1 and 3. By contrast, CPE imposes no such ordering. Since all errors originate from the mediator, a single “noisy” event at the source can result in simultaneous, correlated mistakes by multiple players. It is therefore entirely possible for the likelihood of a multiple-player deviation to exceed that of a single-player one.

**Non-Convexity of CPE.** Example 1 also shows that the set of correlated perfect equilibria is not necessarily convex. In particular, note that while both  $\delta_{(z_1, z_2, y_3)}$  and  $\delta_{(y_1, y_2, y_3)}$  are correlated perfect, any (strict) convex combination is not. To see this, let  $\rho$  denote

<sup>10</sup>For expositional simplicity, we omit the simple normalization needed to form a probability distribution.



	$w_2$	$x_2$	$y_2$	$z_2$		$w_2$	$x_2$	$y_2$	$z_2$
$w_1$	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	$w_1$	0, 0, 0	0, 0, 0	1, 0, 0	1, 0, 0
$x_1$	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	$x_1$	0, 0, 0	0, 0, 0	1, 0, 0	1, 0, 0
$y_1$	0, 1, 0	0, 1, 0	1, -2, 0	-2, 1, 0	$y_1$	0, 0, 0	0, 0, 0	1, -2, 0	-2, 1, 0
$z_1$	0, 1, 0	0, 1, 0	-2, 1, 0	1, -2, 0	$z_1$	0, 0, 0	0, 0, 0	-2, 1, 0	1, -2, 0
	$x_3$					$y_3$			

(a) A three-person game

	$w_2$	$x_2$	$y_2$	$z_2$		$w_2$	$x_2$	$y_2$	$z_2$
$w_1$	0	0	0	0	$w_1$	0	0	1/16	3/16
$x_1$	0	0	0	0	$x_1$	0	0	3/16	1/16
$y_1$	3/16	1/16	0	0	$y_1$	0	0	0	0
$z_1$	1/16	3/16	0	0	$z_1$	0	0	0	0
	$x_3$					$y_3$			

(b) A perfect (direct) correlated equilibrium

FIGURE 2. A perfect direct correlated equilibrium that is not correlated perfect.

such a convex combination. Then

$$S^\rho = \{y_1, z_1\} \times \{y_2, z_2\} \times \{y_3\}.$$

Now consider the deviation plan  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$  defined as follows:

- (i) For  $i = 1, 2$ , set  $\alpha_i(\cdot | s_i) = \delta_{x_i}$  if  $s_i \in S_i^\rho$ ;  $\alpha_i(\cdot | s_i) = \delta_{s_i}$  otherwise.
- (ii) For  $i = 3$ , set  $\alpha_3(\cdot | s_3) = \delta_{s_3}$  for all  $s_3 \in S_3$ .

This plan instructs players 1 and 2 to deviate to  $x_1$  and  $x_2$ , respectively, whenever they are recommended a strategy in  $S_1^\rho$  or  $S_2^\rho$ , while player 3 always follows the recommendation. It is straightforward to check that  $\alpha$  is a dual vector restricted to  $S^\rho$ . However, the plan is collectively profitable: it yields a strict inequality in (4) for every  $s \in \{y_1, z_1\} \times \{y_2, z_2\} \times \{y_3\}$ . By Theorem 1, it follows that  $\rho$  is not correlated perfect.

**Example 2** (A PDCE that is not CPE). Consider a three-person game in Figure 2a where  $S_1 = \{w_1, x_1, y_1, z_1\}$ ,  $S_2 = \{w_2, x_2, y_2, z_2\}$  and  $S_3 = \{x_3, y_3\}$ . For simplicity, player 3's payoffs are normalized to zero. One can verify that the correlated strategy  $\rho$  given in Figure 2b is a PDCE. We defer the details to the appendix (see Appendix B).

We now show that  $\rho$  is not correlated perfect. Since  $S^\rho = S$ , it suffices to exhibit an unrestricted dual vector  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$  that makes (4) strict on some profiles. Define

- (i) for  $i = 1, 2$ :  $\alpha_i(w_i | y_i) = \alpha_i(w_i | z_i) = \alpha_i(w_i | w_i) = \alpha_i(x_i | x_i) = 1$ , and  $\alpha_i(s'_i | s_i) = 0$  otherwise;
- (ii) for  $i = 3$ :  $\alpha_3(x_3 | x_3) = \alpha_3(y_3 | y_3) = 1$ , and  $\alpha_3(s'_3 | s_3) = 0$  otherwise.

By construction, player 3 never deviates, so  $D_3(s, \alpha_3) = 0$  for all  $s$ . For players 1 and 2, the deviation plans send  $y_i$  and  $z_i$  to  $w_i$  while leaving  $w_i$  and  $x_i$  unchanged. Hence, for

any  $s \in \{y_1, z_1\} \times \{y_2, z_2\} \times \{x_3, y_3\}$  we have

$$\sum_{i=1}^3 D_i(s, \alpha_i) = (u_1(w_1, s_{-1}) - u_1(s)) + (u_2(w_2, s_{-2}) - u_2(s)).$$

For example, at  $s = (y_1, y_2, x_3)$ ,

$$D_1(s, \alpha_1) = 0 - 1 = -1, \quad D_2(s, \alpha_2) = 1 - (-2) = +3,$$

so  $\sum_i D_i(s, \alpha_i) = 2 > 0$ .

Similar calculations show that (4) holds with strict inequality for each  $s \in \{y_1, z_1\} \times \{y_2, z_2\} \times \{x_3, y_3\}$ . For  $s \notin \{y_1, z_1\} \times \{y_2, z_2\} \times \{x_3, y_3\}$ , either no deviation is prescribed or the payoff difference is zero, so (4) holds with equality. Thus  $\alpha$  is an unrestricted dual vector that yields a strict inequality on every  $s \in \{y_1, z_1\} \times \{y_2, z_2\} \times \{x_3, y_3\}$ , and by Theorem 1,  $\rho$  is not correlated perfect.

**Why CPE  $\not\subseteq$  PDCE.** In Example 2, we identified a PDCE that is not CPE. The key difference lies in how off-equilibrium beliefs are formed under these two refinements. In a PDCE, trembles are modeled as independent, player-specific mistakes. Each player evaluates her best response conditional on not trembling herself, which introduces private information off the equilibrium path. This allows players to form different—yet internally consistent—stories to rationalize the same off-path event. As we shall see, such disagreements cannot arise under correlated perfection.

In the example, suppose  $w_1$  and  $x_1$  are not played and player 2 is recommended  $y_2$ . From player 2's viewpoint, the most likely event is that player 1 trembled while player 3 followed the equilibrium strategy  $y_3$ . From player 1's perspective, however, when he is recommended  $y_1$ , the same observation is most likely when player 2 trembled and player 3 played the equilibrium strategy  $x_3$ . Thus, players 1 and 2 assign *different* off-path beliefs about player 3 to rationalize the *same* event, and likewise disagree about the identity of the deviating player. By contrast, under CPE, such disagreements are impossible: because there is a single, common source of error—the mediator's mistakes—all players share the same information structure and hence update to a common posterior over off-path profiles.

**Duplication Invariance.** Notice that the first two strategies of players 1 and 2 in Example 2 are payoff-equivalent duplicates. This simplification is purely expositional: one could, for instance, slightly perturb player 3's payoff at  $(w_1, w_2, x_3)$  to make these strategies strategically distinct without changing the argument. Nevertheless, such duplications are essential for sustaining PDCE. As mentioned before, under a PDCE, mistakes are assumed to be independent for each player, conditional on the message they receive. By duplicating strategies, however, the mediator gains the ability to correlate the messages sent to players, thereby inducing correlated mistakes and effectively circumventing the independence assumption. Indeed, as shown in Chatterji and Govindan (2006), in any

finite normal-form game, every PCE—without restricting to the direct mechanism—can be implemented as a PDCE by duplicating players’ pure strategies.

By contrast, our notion of correlated perfection is duplication-invariant: adding or removing duplicate strategies does not affect whether a correlated equilibrium is correlated perfect. This follows directly from the definition of CPE, where mistakes originate from a single mediator whose noisy recommendations treat strategically equivalent actions symmetrically. As a result, such duplications have no effect on the supporting sequence of equilibrium. This property ensures that correlated perfection depends solely on the game’s strategic structure, not its particular representation. We view this as an advantage as it makes the concept robust to distortions caused by duplicated strategies.

## 5. CONCLUSION

We propose a refinement of correlated equilibrium, which we call *correlated perfect equilibrium* (CPE). A CPE is a correlated equilibrium that is robust to arbitrarily small mistakes by the mediator, in contrast to earlier refinements that attribute mistakes to the players. The main benefit of our approach lies in its conceptual simplicity. By tracing the errors to a single, common source, CPE provides a natural framework for correlated mistakes while ensuring belief consistency across players. This distinguishes it from existing player-tremble-based refinements—such as ACE and PDCE—since the perturbation applies to the strategy space that is common to all players. As a result, it does not generate private information beyond what can be inferred from the correlation device. As we illustrate via two examples, there is no set-inclusion relationship between CPE and these existing refinements.

Our main result characterizes the set of CPEs through a dual representation, which reveals its geometric structure—as a finite union of convex polyhedra—and offers a new perspective on strategic stability under noise. The refinement also inherits several desirable properties: like other perfection notions, it eliminates all weakly dominated strategies and is invariant to strategically irrelevant features, such as the duplication of redundant strategies.

## APPENDIX A. OMITTED PROOFS

*Proof of Lemma 1.* Let  $\sigma \in \prod_{i \in N} \Delta S_i$  be a perfect equilibrium of the game  $G$ . By definition (Selten, 1975; Myerson, 1978), there exists a sequence of completely mixed strategy profiles  $(\sigma^k)$  converging to  $\sigma$  such that, for each  $k$ , each player  $i \in N$ , and each  $s_i \in S_i$  with  $\sigma_i(s_i) > 0$ ,

$$\sum_{s_{-i} \in S_{-i}} \sigma_{-i}^k(s_{-i}) u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \sigma_{-i}^k(s_{-i}) u_i(s'_i, s_{-i}) \quad \text{for all } s'_i \in S_i.$$

Now construct a sequence of completely mixed correlated strategies  $(\rho^k)$  by setting, for each  $s \in S$ ,

$$\rho^k(s) := \prod_{i \in N} \sigma_i^k(s_i).$$

Clearly,  $\rho^k$  converges to  $\sigma$ . Moreover, multiplying both sides of the inequality above by  $\sigma_i^k(s_i)$  yields condition (3). Hence every perfect equilibrium is also a correlated perfect equilibrium.  $\square$

Recall that a pure strategy  $s_i$  of player  $i$  is said to be *weakly dominated* if there exists a mixed strategy  $\sigma_i \in \Delta S_i$  such that

$$u_i(s_i, s_{-i}) \leq \sum_{s'_i \in S_i} \sigma_i(s'_i) \cdot u_i(s'_i, s_{-i}), \quad \text{for all } s_{-i} \in S_{-i},$$

and the inequality is strict for at least one  $s_{-i} \in S_{-i}$ . The following claim will be useful in proving Lemma 2.

*Claim 1.* (Myerson, 1991, Theorem 1.7, pp. 30–31). A pure strategy  $s_i \in S_i$  of player  $i$  is weakly dominated if and only if, for every completely mixed correlated strategy  $\rho \in \Delta S$ , the inequality in (3) fails for some  $s'_i \in S_i$ .

*Proof of Lemma 2.* Suppose toward a contradiction that  $\rho$  is a correlated perfect equilibrium and that a weakly dominated strategy  $s_i$  is in its support, i.e.,  $\rho_i(s_i) > 0$ . Since  $\rho_i(s_i) > 0$ , then condition (3) must hold for  $s_i$ . But by Claim 1, for every completely mixed correlated strategy  $\hat{\rho}$ , we have

$$\sum_{s_{-i} \in S_{-i}} \hat{\rho}(s_i, s_{-i}) \cdot u_i(s_i, s_{-i}) < \sum_{s_{-i} \in S_{-i}} \hat{\rho}(s_i, s_{-i}) \cdot u_i(s'_i, s_{-i}) \quad \text{for some } s'_i \in S_i.$$

That is, no sequence of completely mixed correlated strategies  $\rho^k$  converging to  $\rho$  can satisfy (3) for  $s_i$ , a contradiction.  $\square$

To prove Theorem 1, we will use the following version of Farkas' Lemma (see, e.g., Schrijver, 1986, p. 89).

**Lemma 3.** Let  $A$  be an  $m \times n$  matrix and  $b$  be an  $m$ -length vector. Then exactly one of the following holds:

- (i) There exists an  $x \in \mathbb{R}^n$  such that  $Ax \geq b$ .

(ii) There exists a  $y \in \mathbb{R}_+^m$  such that  $y^\top A = 0$  and  $y \cdot b > 0$ .

*Proof of Theorem 1.* We first show the only-if direction by contraposition. Suppose there exists an  $S^\rho$  restricted dual vector  $\alpha$  and a profile  $s'$  such that

$$\sum_{i \in N} \sum_{t_i \in S_i} \alpha_i(t_i | s'_i) (u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})) > 0. \quad (5)$$

Notice that some player  $i$  must be playing a strategy from  $S_i^\rho$  at  $s'$ .

Let  $\rho$  be a correlated equilibrium. Consider an arbitrary sequence of completely mixed correlated strategies  $(\rho^k)$  converging to  $\rho$  as  $k \rightarrow \infty$ . For each profile  $s'$ , multiplying (5) by  $\rho^k(s') > 0$  and then summing over  $s' \in S$  yields

$$\sum_{s' \in S} \rho^k(s') \sum_{i \in N} \sum_{t_i \in S_i} \alpha_i(t_i | s'_i) [u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})] > 0.$$

Rearranging the sums and using that  $\alpha$  is  $S^\rho$ -restricted (so  $\alpha_i(\cdot | s'_i) = \delta_{s'_i}$  whenever  $s'_i \notin S_i^\rho$ ), we obtain

$$\sum_{i \in N} \sum_{s'_i \in S_i^\rho} \sum_{t_i \in S_i} \alpha_i(t_i | s'_i) \sum_{s'_{-i} \in S_{-i}} \rho^k(s'_i, s'_{-i}) [u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})] > 0.$$

But this implies that there exist some  $i$ ,  $s'_i \in S_i^\rho$  and  $t_i \in S_i$  such that

$$\sum_{s'_{-i} \in S_{-i}} \rho^k(s'_i, s'_{-i}) [u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})] > 0,$$

which contradicts (3). As the choice of  $\rho^k$  was arbitrary, this shows that every completely mixed sequence converges to  $\rho$  violates (3). Thus, no supporting sequence exists and we conclude that  $\rho$  is not correlated perfect.

We now show the if-direction. Consider a  $(\sum_{i \in N} |S_i^\rho| \cdot |S_i| + |S|) \times |S|$  matrix  $A$  where

- (i) Each column corresponds to a strategy profile  $s' \in S$ .
- (ii) The first block of rows is indexed by triples  $(i, \tilde{s}_i, s_i)$  with  $\tilde{s}_i \in S_i^\rho$  and  $s_i \in S_i$ .  
The entry at row  $(i, \tilde{s}_i, s_i)$  and column  $s' = (s'_i, s'_{-i})$  is

$$A[(i, \tilde{s}_i, s_i), s'] = \begin{cases} u_i(\tilde{s}_i, s'_{-i}) - u_i(s_i, s'_{-i}), & \text{if } s'_i = \tilde{s}_i, \\ 0, & \text{otherwise.} \end{cases}$$

- (iii) The last block of  $|S|$  rows consists of the  $|S| \times |S|$  identity matrix.

Let  $b \in \mathbb{R}^{\sum_{i \in N} |S_i^\rho| \cdot |S_i| + |S|}$  be a vector where the first  $\sum_{i \in N} |S_i^\rho| \cdot |S_i|$  entries are 0 and the last  $|S|$  entries are 1.

Let  $\rho$  be a correlated equilibrium. Suppose for every dual vector  $\alpha = (\alpha_i)_{i \in N}$  that is restricted to  $S^\rho$  and every  $s' \in S$ , (4) holds with equality:

$$\sum_{i \in N} D_i(s', \alpha_i) = \sum_{i \in N} \sum_{t_i \in S_i} \alpha_i(t_i | s'_i) (u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})) = 0.$$

We will argue first that condition (ii) in Lemma 3 is impossible and thus condition (i) must hold. We then use condition (i) to show that  $\rho$  is a CPE.

To this end, assume towards a contradiction that there exists  $y = (\alpha, \beta) \in \mathbb{R}_+^{\sum_i |S_i^\rho| \cdot |S_i| + |S|}$  with  $y^\top A = 0$  and  $y \cdot b > 0$  where (i)  $\alpha = (\alpha_i)_{i \in N}$  and each  $\alpha_i(\tilde{s}_i | s_i) \geq 0$  is associated with the row  $(i, \tilde{s}_i, s_i)$  and (ii)  $\beta(s') \geq 0$  is associated with the identity row for column  $s'$ .

For any  $s' \in S$ , the  $s'$ -th component of  $y^\top A$  is

$$(y^\top A)_{s'} = \sum_{i \in N} \sum_{s_i \in S_i} \alpha_i(s'_i | s_i) [u_i(s'_i, s'_{-i}) - u_i(s_i, s'_{-i})] + \beta(s') = 0, \quad (6)$$

since the entries are zero in  $A$  unless  $\tilde{s}_i = s'_i$  and so the inner sum over  $\tilde{s}_i$  collapses to  $s'_i$ .

For each  $(i, s'_i)$  where  $s'_i \in S_i^\rho$ , set  $d_i(s'_i) := \sum_{t_i \in S_i} \alpha_i(s'_i | t_i) \geq 0$  and define

$$\hat{\alpha}_i(t_i | s'_i) := \begin{cases} \alpha_i(s'_i | t_i) / d_i(s'_i), & d_i(s'_i) > 0, \\ \mathbf{1}(t_i = s'_i), & d_i(s'_i) = 0. \end{cases}$$

Then we can write

$$\sum_{s_i \in S_i} \alpha_i(s'_i | s_i) [u_i(s'_i, s'_{-i}) - u_i(s_i, s'_{-i})] = d_i(s'_i) \sum_{t_i \in S_i} \hat{\alpha}_i(t_i | s'_i) [u_i(s'_i, s'_{-i}) - u_i(t_i, s'_{-i})]. \quad (7)$$

Recall that an  $S^\rho$ -restricted dual vector  $\alpha$  satisfies: for each  $i$  and  $s_i \in S_i$ ,  $\alpha_i(\cdot | s_i)$  is a probability distribution, and  $\alpha_i(\cdot | s_i) = \delta_{s_i}$  whenever  $s_i \notin S_i^\rho$ . Fix  $i$  and construct an  $S^\rho$ -restricted dual vector  $\tilde{\alpha}^{(i)}$  by

$$\tilde{\alpha}_k^{(i)}(\cdot | s_k) = \begin{cases} \hat{\alpha}_i(\cdot | s'_i), & k = i \text{ and } s_k = s'_i, \\ \delta_{s_k}, & \text{otherwise.} \end{cases}$$

Applying the hypothesis at  $s'$  with  $\tilde{\alpha}^{(i)}$  (the  $k \neq i$  terms vanish because  $\delta_{s'_k}$  gives zero gain) gives

$$\sum_{t_i} \hat{\alpha}_i(t_i | s'_i) [u_i(t_i, s'_{-i}) - u_i(s'_i, s'_{-i})] = 0.$$

It then follows from (7) that for each  $i$  and  $s'_i \in S_i^\rho$ ,

$$\sum_{s_i \in S_i} \alpha_i(s'_i | s_i) [u_i(s'_i, s'_{-i}) - u_i(s_i, s'_{-i})] = 0.$$

Summing over  $i$  yields that the first term in (6) equals zero and this holds for every  $s'$ . Therefore,  $y \cdot b = \sum_{s' \in S} \beta(s') = 0$ , contradicting  $y \cdot b > 0$ . Thus, (ii) in Lemma 3 cannot hold, and so the primal system  $Ax \geq b$  must be feasible. That is, there exists  $\mu \in \mathbb{R}^{|S|}$  such that

$$\begin{aligned} \sum_{s_{-i} \in S_{-i}} \mu(\tilde{s}_i, s_{-i}) [u_i(\tilde{s}_i, s_{-i}) - u_i(s_i, s_{-i})] &\geq 0 & \forall i \in N, \tilde{s}_i \in S_i^\rho, s_i \in S_i, \\ \mu(s) &\geq 1 & \forall s \in S. \end{aligned} \quad (8)$$

Now, we construct a supporting sequence for  $\rho$  to be a CPE. For each  $k \in \mathbb{N}$ , define

$$\hat{\rho}^k(s) := \rho(s) + \frac{1}{k} \cdot \mu(s)$$

and normalize to obtain

$$\rho^k(s) := \frac{\hat{\rho}^k(s)}{\sum_{\tau \in S} \hat{\rho}^k(\tau)}.$$

Then  $\rho^k$  is completely mixed since  $\mu(s) \geq 1$  and  $\rho^k \rightarrow \rho$  as  $k \rightarrow \infty$ . It remains to show that (3) holds for  $\rho^k$ . Fix any  $i$ ,  $s_i \in S_i^\rho$  and any deviation  $s'_i \in S_i$ , we can write

$$\begin{aligned} & \sum_{s_{-i} \in S_{-i}} \rho^k(s_i, s_{-i}) (u_i(s_i, s_{-i}) - u_i(s'_i, s_{-i})) \\ &= \frac{1}{\sum_{s \in S} \hat{\rho}^k(s)} \left( \sum_{s_{-i} \in S_{-i}} \rho(s) (u_i(s_i, s_{-i}) - u_i(s'_i, s_{-i})) + \frac{1}{k} \sum_{s_{-i} \in S_{-i}} \mu(s) (u_i(s_i, s_{-i}) - u_i(s'_i, s_{-i})) \right). \end{aligned}$$

The first term is nonnegative since  $\rho$  is a correlated equilibrium and the second term is nonnegative by (8). Hence, condition (3) holds along the sequence  $\rho^k$  and thus we conclude that  $\rho$  is a correlated perfect equilibrium.  $\square$

## APPENDIX B. OMITTED DETAILS IN EXAMPLE 2

For completeness, we recall the definition of a perfect direct correlated equilibrium from [Dhillon and Mertens \(1996\)](#). For  $k \in \mathbb{N}$ , let  $\sigma_i^k(s'_i | s_i)$  denote the probability that player  $i$  trembles to  $s'_i$  conditional on being recommended  $s_i$ , and let  $\sigma^k = (\sigma_i^k(\cdot | s_i))_{s_i \in S_i, i \in N}$  be a profile of conditional strategies. For  $\rho \in \Delta S$  and  $s \in S$ , define

$$\rho_{\sigma^k}^i(s_i, s_{-i}) = \sum_{s'_{-i} \in S_{-i}} \rho(s_i, s'_{-i}) \prod_{j \neq i} \sigma_j^k(s_j | s'_j),$$

which represents, from player  $i$ 's perspective, the probability that  $s$  is played when the mediator draws from  $\rho$  and the other players independently tremble according to  $\sigma^k$ . A distribution  $\rho \in \Delta S$  is a *perfect direct correlated equilibrium* (PDCE) if there exists a sequence of completely mixed conditional strategy profiles  $(\sigma^k)$  such that:

- (i) For each  $i \in N$  and  $s_i \in S_i$ ,  $\sigma_i^k(s_i | s_i) \rightarrow 1$  as  $k \rightarrow \infty$ ; and
- (ii) For every  $i \in N$ , and  $s_i \in S_i$ ,

$$\sum_{s_{-i} \in S_{-i}} \rho_{\sigma^k}^i(s_i, s_{-i}) \cdot u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \rho_{\sigma^k}^i(s_i, s_{-i}) \cdot u_i(s'_i, s_{-i}), \quad \text{for all } s'_i \in S_i, k \in \mathbb{N}.$$

That is, a PDCE is a correlated equilibrium in which every player  $i$ 's recommendation remains optimal against small independent trembles of the opponents, conditional on  $i$  not trembling.

Now, to see that the correlated equilibrium  $\rho$  in Figure 2b is a PDCE, consider the following sequence of independent trembles. Throughout the rest of this section, we use  $\varepsilon > 0$  as a shorthand for  $\varepsilon = 1/k$  where  $k \in \mathbb{N}$  indexes the supporting sequence introduced above. For each sufficiently small  $\varepsilon > 0$ , define



(i) For  $i = 1, 2$ ,

$$\sigma_i^\varepsilon(s'_i | s_i) = \begin{cases} 1 - 2\varepsilon - \varepsilon^2, & \text{if } s'_i = s_i \in \{w_i, x_i\}, \\ 1 - 3\varepsilon, & \text{if } s'_i = s_i \in \{y_i, z_i\}, \\ \varepsilon^2, & \text{if } (s'_i, s_i) \in \{(z_i, w_i), (y_i, x_i)\}, \\ \varepsilon, & \text{otherwise.} \end{cases}$$

(ii) For  $i = 3$ ,

$$\sigma_3^\varepsilon(s'_3 | s_3) = \begin{cases} 1 - \varepsilon, & s'_3 = s_3, \\ \varepsilon, & s'_3 \neq s_3. \end{cases}$$

Condition (i) for PDCE is clearly satisfied. For condition (ii), consider player 1 and suppose the recommendation is  $w_1$ . Under  $\rho$ , the only opponent recommendations with positive probability are  $(y_2, y_3)$  and  $(z_2, y_3)$ . Hence, for any opponent profile  $(s_2, s_3)$ ,

$$\rho_{\sigma^\varepsilon}^1(w_1, s_2, s_3) = \frac{1}{16} \cdot \sigma_2^\varepsilon(s_2 | y_2) \cdot \sigma_3^\varepsilon(s_3 | y_3) + \frac{3}{16} \cdot \sigma_2^\varepsilon(s_2 | z_2) \cdot \sigma_3^\varepsilon(s_3 | y_3).$$

From player 1's perspective, the game in Figure 2a would be played according to

	$w_2$	$x_2$	$y_2$	$z_2$		$w_2$	$x_2$	$y_2$	$z_2$
	$\frac{\varepsilon^2}{4}$	$\frac{\varepsilon^2}{4}$	$\frac{\varepsilon}{16}$	$\frac{3\varepsilon}{16} - \frac{\varepsilon^2}{2}$		$\frac{\varepsilon - \varepsilon^2}{4}$	$\frac{\varepsilon - \varepsilon^2}{4}$	$\frac{1 - \varepsilon}{16}$	$\frac{3 - 11\varepsilon + 8\varepsilon^2}{16}$
$w_1$					$w_1$				
$x_1$	0	0	0	0	$x_1$	0	0	0	0
$y_1$	0	0	0	0	$y_1$	0	0	0	0
$z_1$	0	0	0	0	$z_1$	0	0	0	0
	$x_3$					$y_3$			

The expected gain from deviating to  $y_1$  is

$$\rho_{\sigma^\varepsilon}^1(w_1, y_2, x_3) - 2\rho_{\sigma^\varepsilon}^1(w_1, z_2, x_3) - 3\rho_{\sigma^\varepsilon}^1(w_1, z_2, y_3) = -\frac{9}{16} + \frac{7}{4}\varepsilon - \frac{1}{2}\varepsilon^2.$$

Likewise, the expected gain from deviating to  $z_1$  is

$$-2\rho_{\sigma^\varepsilon}^1(w_1, y_2, x_3) + \rho_{\sigma^\varepsilon}^1(w_1, z_2, x_3) - 3\rho_{\sigma^\varepsilon}^1(w_1, y_2, y_3) = -\frac{3}{16} + \frac{1}{4}\varepsilon - \frac{1}{2}\varepsilon^2.$$

Both expressions are negative for  $\varepsilon$  small enough. The expected gain from deviating to  $x_1$  is zero since  $w_1$  and  $x_1$  are payoff-equivalent for player 1. Hence, no deviation is profitable for player 1 when recommended  $w_1$ . By analogous arguments, one can verify that player 1 has no incentive to deviate when recommended  $x_1$ ,  $y_1$ , or  $z_1$ . By symmetry, the same arguments also hold for player 2. Finally, since player 3's payoffs are constant across all outcomes, any recommendation is trivially optimal. Therefore, these trembles satisfy condition (ii) in the definition of PDCE, and we conclude that the distribution shown in Figure 2b is a perfect direct correlated equilibrium.

## REFERENCES

- AUMANN, R. J. (1974): “Subjectivity and Correlation in Randomized Strategies,” *Journal of Mathematical Economics*, 1, 67–96.
- (1987): “Correlated Equilibrium as an Expression of Bayesian Rationality,” *Econometrica*, 55, 1–18.
- CHATTERJI, S. AND S. GOVINDAN (2006): “Message Spaces for Perfect Correlated Equilibria,” *Economic Theory*, 28, 475–479.
- DASKALAKIS, C., P. W. GOLDBERG, AND C. H. PAPADIMITRIOU (2009): “The complexity of computing a Nash equilibrium,” *Communications of the ACM*, 52, 89–97.
- DHILLON, A. AND J. F. MERTENS (1996): “Perfect Correlated Equilibria,” *Journal of Economic Theory*, 68, 279–302.
- GRANT, S. AND R. STAUBER (2022): “Delegation and ambiguity in correlated equilibrium,” *Games and Economic Behavior*, 132, 487–509.
- HART, S. AND D. SCHMEIDLER (1989): “Existence of correlated equilibria,” *Mathematics of Operations Research*, 14, 18–25.
- LUO, X., Y. QIAO, AND Y. SUN (2022): “A revelation principle for correlated equilibrium under trembling-hand perfection,” *Journal of Economic Theory*, 200, 105396.
- MOULIN, H. AND J.-P. VIAL (1978): “Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon,” *International Journal of Game Theory*, 7, 201–221.
- MYERSON, R. B. (1978): “Refinements of the Nash Equilibrium Concept,” *International Journal of Game Theory*, 7, 73–80.
- (1986): “Acceptable and Predominant Correlated Equilibria,” *International Journal of Game Theory*, 15, 133–154.
- (1991): *Game Theory: Analysis of Conflict*, Harvard University Press.
- NASH, J. F. (1950): “Non-cooperative games,” .
- NAU, R. F. AND K. F. MCCARDLE (1990): “Coherent Behavior in Noncooperative Games,” *Journal of Economic Theory*, 50, 424–444.
- PAPADIMITRIOU, C. H. AND T. ROUGHGARDEN (2008): “Computing correlated equilibria in multi-player games,” *Journal of the ACM (JACM)*, 55, 1–29.
- SCHRIJVER, A. (1986): *Theory of Linear and Integer Programming*, Chichester; New York: Wiley.
- SELTEN, R. (1975): “Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games,” *International Journal of Game Theory*, 4, 25–55.