

Locally Robust Implementation of Efficient Bilateral Trade with Correlated Beliefs*

Takashi Kunimoto[†] Cuiling Zhang[‡]

This Version: July 2026

Abstract

We identify the ex ante welfare (EAW) condition as a necessary requirement to implement ex post efficient bilateral trade in any finite type space with interdependent values and correlated beliefs. As these finite settings become finer to approximate a continuous type space, we derive a limit EAW condition by taking the EAW condition in finite settings to its limit. We show that this limit condition trivially holds in the benchmark continuous setting admitting a full-support density function. We then insist on locally robust implementation by requiring efficient trade to be implemented uniformly across all finite type spaces that approximate the benchmark continuous type space. Our main result shows that under high interdependence, locally robust implementation of efficient trade is impossible. We thus show that the negative results of Myerson and Satterthwaite (1983) and Fieseler, Kittsteiner, and Moldovanu (2003) under independent beliefs can also emerge as the robust limit of discrete environments even when beliefs are correlated.

JEL Classification: C72, D78, D82.

Keywords: bilateral trade, the ex ante welfare condition, interdependence, correlation.

*This work expands upon preliminary material previously developed by Kunimoto and Zhang (2024). We thank Shurojit Chatterji, Seungjin Han, Bobby Pakzad-Hurson, Rene Saran, Sergei Severinov, Bruno Strulovici, Takuro Yamashita, and the audience at SSCW2026 at University of Tokyo for useful comments.

[†]School of Economics, Singapore Management University; tkunimoto@smu.edu.sg

[‡]Institute for Advanced Economic Research, Dongbei University of Finance and Economics; cuilingzhang01@outlook.com

1 Introduction

This paper characterizes when efficient, voluntary bilateral trade can be achieved in environments with interdependent values and correlated beliefs. We model a single buyer and seller with two-sided asymmetric information, where each agent’s private information can affect both valuations, and their beliefs about the opponent’s valuations may be correlated.

By the revelation principle, implementing efficient, voluntary trade reduces to finding a *direct revelation* mechanism where truth-telling is a Bayesian Nash equilibrium. Specifically, the mechanism must satisfy *Bayesian incentive compatibility (BIC)*, *decision efficiency (EFF)*, *interim individual rationality (IIR)*, and *ex post budget balance (BB)*. We define the class of mechanisms satisfying these four properties as *desired* mechanisms.

The bilateral trade literature has focused primarily on two distinct informational environments: continuous type spaces with independent beliefs, and finite type spaces with either independent or correlated beliefs.¹

Within the continuous framework, Myerson and Satterthwaite (1983) establish the non-existence of desired mechanisms in private value environments, an impossibility result that Gresik (1991b) and Fieseler, Kittsteiner, and Moldovanu (2003, Theorem 5) (henceforth, FKM) extend to interdependent values. By contrast, in a two-type setting with independent beliefs and private values, Matsuo (1989) overturn Myerson and Satterthwaite’s (1983) impossibility result by identifying a restrictive but still permissive condition for the existence of desired mechanisms. In the same environment, Gresik (1991a) accommodates correlated beliefs and establishes a corresponding condition for the existence of desired mechanisms. Thus, Matsuo (1989) and Gresik (1991a) highlight how the finite type structure can sometimes restore possibility.

To overcome the restrictiveness of the two-type paradigm, Kosenok and Severinov (2008) examine a general mechanism design framework – encompassing bilateral trade – with finite types, correlated beliefs, and interdependent values. Their Theorem 1 establishes that the Crémer-McLean condition – due to Crémer-McLean (1985, 1988) – and the identifiability condition together guarantee the existence of desired mechanisms.² However, Kosenok and Severinov (2008, p.135) demonstrate that when the environment is restricted to two agents, as in bilateral trade, the identifiability condition holds only under independent beliefs. This restriction

¹While the literature on bilateral trade is vast, we focus our attention on the works most closely related to our paper.

²The reader is referred to Kosenok and Severinov (2008) for the definitions of these conditions.

makes it impossible to apply Kosenok and Severinov’s (2008) existence result to the analysis of bilateral trade with correlated beliefs.³

We summarize and categorize the key differences between these results we discussed above in the following table.

	Model	Type Space	Beliefs	Valuations
MS (1983)	Bilateral Trade	Continuous	Independent	Private
Matsuo (1989)	Bilateral Trade	Two-Type	Independent	Private
Gresik (1991a)	Bilateral Trade	Two-Type	Correlated	Private
Gresik (1991b, Theorem 3)	Bilateral Trade	Continuous	Independent	Interdependent
FKM (2003, Theorem 5)	Bilateral Trade	Continuous	Independent	Interdependent
KS (2008)	General	Finite	Correlated	Interdependent
This paper	Bilateral Trade	Finite	Correlated	Interdependent

Table 1: Comparison of Related Work. MS stands for Myerson and Satterthwaite; FKM stands for Fieseler, Kittsteiner, and Moldovanu; KS stands for Kosenok and Severinov; “General” stands for general mechanism design problems encompassing bilateral trade; “Continuous” stands for the one-dimensional continuous type space; “Finite” stands for the general finite type space; and “Two-Type” stands for the special case of finite type spaces where each agent has only two types.

The existing literature points to an intriguing asymmetry: while correlated beliefs can facilitate the existence of desired mechanisms in finite type spaces (e.g., Gresik (1991a) and Kosenok and Severinov (2008)), interdependent valuations appear to hinder the existence of desired mechanisms in continuous type spaces (e.g., Gresik (1991b) and FKM (2003)). This tension motivates the central questions of this paper: First, under what conditions does a desired mechanism exist in a bilateral trade environment with finite types, correlated beliefs, and interdependent values? Second, whether and how correlated beliefs can reconcile the dichotomy between positive results in discrete environments and negative results in continuous environments.

In our model, the buyer and the seller each possess a finite type space of size $m \geq 2$ and types are ordered monotonically. Throughout the paper, we require trades to be always ex post efficient, meaning trade occurs if and only if the buyer’s valuation is at least as high as the seller’s. We assume the type space satisfies the *single-crossing property* and that the efficient allocation rule is *responsive*.

³Garratt and Pycia (2025) recently examine how relaxing the assumption of quasilinearity changes the Myerson and Satterthwaite’s (1983) impossibility result. They show that, in a continuous type space with private values and either independent or correlated beliefs, efficient trade is generically possible under any departure from quasilinearity, as long as the asymmetry of information is not too large (Theorem 2). This is in contrast with our analysis where we maintain the quasilinearity and allow for interdependent values.

Responsiveness rules out cases where any two distinct types of an agent always induce the identical decision outcome. Proposition 1 demonstrates that these two properties uniquely determine the efficient allocation rule x^* , as detailed in Table 2. The rest of our analysis focuses on this rule x^* .

To identify our key condition, we require all mechanisms to satisfy the *no-trade-then-no-payments* (NTNP) property, which dictates that no monetary transfers are made if trade does not occur. We introduce the *ex ante welfare* (EAW) condition and show it to be necessary for the existence of a desired mechanism in any finite type space (Theorem 1). Intuitively, the EAW condition requires that the ex ante gains from trade be sufficiently large to deter agents from deviating from truth-telling.

Section 5 explores the implications of efficient bilateral trade in a continuous type space. We consider a specialized setting where each agent’s type space is a finite grid over $[0, 1]$ and valuation functions are symmetric and linear. By taking the limit of the EAW condition as the finite type space converges to the benchmark continuous type space, we derive a limit EAW condition. This limit condition holds trivially in standard continuous settings with a full-support density (Lemma 1). To obtain a meaningful implication in a continuous type space, we propose *locally robust implementation* – meaning efficient trade must be implementable uniformly across all finite type spaces that approximate the benchmark continuous type space. Then, Theorem 2 shows that under high interdependence, the EAW condition fails in a finite type space that approximates the benchmark continuous type space. Thus, high preference interdependence renders locally robust implementation impossible. This finding provides a stark contrast to the existing literature: while Myerson and Satterthwaite (1983) and FKM (2003, Theorem 5) establish impossibility under independent beliefs, we show that, if the degree of interdependence is sufficiently high, these negative results emerge as the robust limit of discrete environments even when beliefs are correlated.

In Section 6, we revisit Theorem 5 of FKM (2003) and apply it to the benchmark continuous type space introduced in Section 5.3, where we analyze the locally robust implementation of efficient bilateral trade. We explicitly show that their identified condition for the existence of desired mechanisms is violated in our setting, thereby establishing the impossibility of efficient bilateral trade in environments with interdependent values and independent beliefs.

The rest of the paper is organized as follows. Section 2 introduces the general notation and basic concepts. In Section 3, we impose the single-crossing property and responsiveness to uniquely pin down the efficient decision rule x^* described in

Table 2. Section 4 introduces the no-trade-then-no-payments (NTNP) property, identifies the ex ante welfare (EAW) condition, and shows that the EAW condition is necessary for the existence of desired mechanisms satisfying NTNP. Section 5 leverages the notion of locally robust implementation to explore the implications of efficient bilateral trade in a stylized continuous type space with symmetric, linear valuations. In Section 6, we discuss the relationship between our paper and Theorem 5 of FKM (2003). Finally, Section 7 concludes. All the proofs omitted from the main text appear in the Appendix.

2 Preliminaries

A seller (agent 1) has one indivisible object for sale and there is one potential buyer (agent 2). Each agent has some private information concerning the value of the good, which is summarized as their own type θ_i . In what follows, we assume that each agent's type space is one dimensional and that both agents have the same number of types denoted by m . For each agent $i \in \{1, 2\}$, their type space is $\Theta_i = \{\theta_{i1}, \dots, \theta_{im}\}$ with $0 < \theta_{i1} < \dots < \theta_{im}$. We denote by θ_{ik} a generic element of agent i 's type space. Throughout the paper, we use the notation convention that $\Theta = \Theta_1 \times \Theta_2$; and $\Theta_{-i} = \Theta_j$ where $j \neq i$ with a generic element $\theta_{j\ell}$.

Each agent's valuation for the good depends not only on his own types, but also on the other's types. For each agent $i \in \{1, 2\}$ and type profile $(\theta_{ik}, \theta_{j\ell}) \in \Theta$ with $j \neq i$, we denote by $v_i(\theta_{ik}, \theta_{j\ell}) \geq 0$ their valuation function of the good. Hence, we consider an interdependent values environment. We assume that for each agent i , $v_i(\theta_{ik}, \theta_{j\ell})$ is strictly increasing in agent i 's own type θ_{ik} and nondecreasing in the other's type $\theta_{j\ell}$. We further impose the single crossing condition on the valuation functions.

Definition 1. The valuation functions satisfy the *single crossing condition* if for each agent $i \in \{1, 2\}$, each $k, k' \in \{1, \dots, m\}$ with $k \neq k'$, and each $\ell \in \{1, \dots, m\}$,

$$\frac{v_i(\theta_{ik'}, \theta_{j\ell}) - v_i(\theta_{ik}, \theta_{j\ell})}{\theta_{ik'} - \theta_{ik}} > \frac{v_j(\theta_{ik'}, \theta_{j\ell}) - v_j(\theta_{ik}, \theta_{j\ell})}{\theta_{ik'} - \theta_{ik}},$$

where $j \neq i$.

Types are drawn from a full-support probability distribution μ over Θ : $\mu(\theta_{1k}, \theta_{2\ell}) > 0$ for each type profile $(\theta_{1k}, \theta_{2\ell}) \in \Theta$ and $\sum_{k=1}^m \sum_{\ell=1}^m \mu(\theta_{1k}, \theta_{2\ell}) = 1$. For every $k \in \{1, \dots, m\}$, we denote by $\mu_i(\theta_{ik}) = \sum_{\ell=1}^m \mu(\theta_{ik}, \theta_{j\ell})$ the marginal probability that agent i is of type θ_{ik} . We further define the following notation for conditional probabilities: for each type profile $(\theta_{1k}, \theta_{2\ell}) \in \Theta$, $\mu(\theta_{2\ell}|\theta_{1k}) \equiv \mu(\theta_{1k}, \theta_{2\ell})/\mu_1(\theta_{1k})$

and $\mu(\theta_{1k}|\theta_{2\ell}) \equiv \mu(\theta_{1k}, \theta_{2\ell})/\mu_2(\theta_{2\ell})$. (There is some abuse of notation in using the same symbol μ for probability distributions over different spaces, but this leads to no confusion.)

Let $q_i \in Q = [0, 1]$ denote the probability that agent i is allocated the good. Preferences of each agent depends on the probability of being allocated the good, the type profile and their monetary transfer. Each agents i 's utility function $u_i : Q \times \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ is represented by

$$u_i(q_i, \theta, p_i) = q_i v_i(\theta_{ik}, \theta_{j\ell}) + p_i$$

where $q_1 + q_2 = 1$, $j \neq i$, and p_i denotes the monetary transfer received by agent i .

A *direct mechanism* is defined as a triplet $\Gamma = (\Theta, x, t)$ where Θ_1 and Θ_2 are the set of actions available to the seller and buyer, respectively; $x : \Theta \rightarrow [0, 1]$ is the *decision rule* which specifies the probability that trade occurs; and $(t_1, t_2) : \Theta \rightarrow \mathbb{R}^2$ is the *transfer rule* which describes the monetary transfers to both agents. By the revelation principle, we lose nothing to focus on direct mechanisms. In what follows, we simply call them mechanisms. We shall now discuss the properties we want a mechanism to satisfy.

Definition 2. A mechanism (x, t) satisfies *Bayesian incentive compatibility* (BIC) if, for all agent $i \in \{1, 2\}$ and all $k, k' \in \{1, \dots, m\}$,

$$\begin{aligned} & \sum_{\ell=1}^m \mu(\theta_{j\ell}|\theta_{ik}) [x_i(\theta_{ik}, \theta_{j\ell}) v_i(\theta_{ik}, \theta_{j\ell}) + t_i(\theta_{ik}, \theta_{j\ell})] \\ & \geq \sum_{\ell=1}^m \mu(\theta_{j\ell}|\theta_{ik'}) [x_i(\theta_{ik'}, \theta_{j\ell}) v_i(\theta_{ik}, \theta_{j\ell}) + t_i(\theta_{ik'}, \theta_{j\ell})], \end{aligned}$$

where $j \neq i$, and $x_i(\theta_{ik}, \theta_{j\ell})$ denotes the probability that agent i is allocated the good given the type report $(\theta_{ik}, \theta_{j\ell})$.

We also assume that each agent has the option of not participating in the mechanism (Θ, x, t) and let $U_i^O(\theta_{ik})$ be the expected utility of agent i with type θ_{ik} from non-participation. To be specific,

$$U_1^O(\theta_{1k}) = \sum_{\ell=1}^m \mu(\theta_{2\ell}|\theta_{1k}) v_1(\theta_{1k}, \theta_{2\ell}) \text{ for all } k \in \{1, \dots, m\}$$

and

$$U_2^O(\theta_{2\ell}) = 0 \text{ for all } \ell \in \{1, \dots, m\}.$$

Definition 3. A mechanism (x, t) satisfies *interim individual rationality* (IIR) if, for each agent $i \in \{1, 2\}$ and each $k \in \{1, \dots, m\}$,

$$\sum_{\ell=1}^m \mu(\theta_{j\ell} | \theta_{ik}) [x_i(\theta_{ik}, \theta_{j\ell}) v_i(\theta_{ik}, \theta_{j\ell}) + t_i(\theta_{ik}, \theta_{j\ell})] \geq U_i^O(\theta_{ik}),$$

where $j \neq i$.

See, for example, Myerson and Satterthwaite (1983), Matsuo (1989), and Gresik (1991a, 1991b) for the same treatment of IIR. Throughout the paper, we impose the following decision efficiency of bilateral trade.

Definition 4. A mechanism (x, t) satisfies *decision efficiency* (EFF) if, for each $(\theta_{1k}, \theta_{2\ell}) \in \Theta$,

$$x(\theta_{1k}, \theta_{2\ell}) = \begin{cases} 1 & \text{if } v_2(\theta_{1k}, \theta_{2\ell}) \geq v_1(\theta_{1k}, \theta_{2\ell}) \\ 0 & \text{if } v_2(\theta_{1k}, \theta_{2\ell}) < v_1(\theta_{1k}, \theta_{2\ell}), \end{cases}$$

where $x(\theta_{1k}, \theta_{2\ell})$ denotes the probability of trade given the type report $(\theta_{1k}, \theta_{2\ell})$.

We denote by x^* the efficient decision rule. In our paper, we further impose *ex post* budget balance.

Definition 5. A mechanism (x, t) satisfies *ex post budget balance* (BB) if, for each $(\theta_{1k}, \theta_{2\ell}) \in \Theta$,

$$t_1(\theta_{1k}, \theta_{2\ell}) + t_2(\theta_{1k}, \theta_{2\ell}) = 0.$$

In the rest of the paper, we refer to mechanisms satisfying BIC, IIR, EFF, and BB simply as *desired* mechanisms.

3 The Efficient Decision Rule

There are many different efficient decision rules we can consider in a finite type space. To make our analysis tractable, we restrict the set of efficient decision rules to a small class of efficient decision rules. In particular, we focus on what we call *responsive* decision rules.⁴

Definition 6. An efficient decision rule x^* is *responsive* if the following two requirements are satisfied:

⁴The terminology is adopted from Bergemann and Morris (2009, p.1187), who propose responsiveness as a one-to-one property of social choice functions in their analysis of robust implementation.

1. for any two distinct types of the seller, $\theta_{1k}, \theta_{1k'} \in \Theta_1$, there exists some $\theta_{2\ell} \in \Theta_2$ such that

$$x^*(\theta_{1k}, \theta_{2\ell}) \neq x^*(\theta_{1k'}, \theta_{2\ell});$$

2. for any two distinct types of the buyer, $\theta_{2\ell}, \theta_{2\ell'} \in \Theta_2$, there exists some $\theta_{1k} \in \Theta_1$ such that

$$x^*(\theta_{1k}, \theta_{2\ell}) \neq x^*(\theta_{1k}, \theta_{2\ell'}).$$

We show below that responsiveness together with the single crossing condition pins down the unique efficient decision rule in any finite type space.

Proposition 1. Suppose that the single crossing condition holds; the efficient decision rule x^* is responsive; and $x^*(\theta_{11}, \theta_{21}) = 1$.⁵ Then, the efficient decision rule is uniquely determined as follows: for each $(\theta_{1k}, \theta_{2\ell}) \in \Theta$,

$$x^*(\theta_{1k}, \theta_{2\ell}) = \begin{cases} 1 & \text{if } k \leq \ell \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The proof is in the Appendix. □

The efficient decision rule is depicted in the table below. Henceforth, our analysis will be restricted to this specific rule.

$x^*(\cdot)$	θ_{21}	θ_{22}	θ_{23}	$\cdots$	θ_{2m}
θ_{11}	1	1	1	$\cdots$	1
θ_{12}	0	1	1	$\cdots$	1
θ_{13}	0	0	1	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
θ_{1m}	0	0	0	$\cdots$	1

Table 2: The Efficient Decision Rule $x^*(\cdot)$

4 The Ex Ante Welfare (EAW) Condition

We further impose a property on the class of mechanisms: if no trade occurs, no monetary transfers are made.⁶

⁵If, on the other hand, $x^*(\theta_{11}, \theta_{21}) = 0$, we can similarly show that the efficient decision rule is uniquely determined: for each $(\theta_{1k}, \theta_{2\ell}) \in \Theta$, $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ if and only if $k \leq \ell - 1$. Thus, it is essentially without loss of generality to assume $x^*(\theta_{11}, \theta_{21}) = 1$.

⁶Theorem 1 of Gresik (1991a) also imposes this property.

Definition 7. A transfer rule $t : \Theta \rightarrow \mathbb{R}^2$ satisfies the *no-trade-then-no-payments* (NTNP) property if, for any $(\theta_{1k}, \theta_{2\ell}) \in \Theta$,

$$t_1(\theta_{1k}, \theta_{2\ell}) = t_2(\theta_{1k}, \theta_{2\ell}) = 0 \text{ whenever } x(\theta_{1k}, \theta_{2\ell}) = 0.$$

We are ready to introduce the *ex ante welfare condition*, which will be shown to be a necessary condition for the existence of desired mechanisms satisfying NTNP.

Definition 8. The distribution μ and the collection of valuations $(v_1(\theta), v_2(\theta))_{\theta \in \Theta}$ satisfy the *ex ante welfare (EAW) condition* if the following inequality holds:

$$\begin{aligned} & \sum_{k=1}^m \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) (v_2(\theta_{1k}, \theta_{2\ell}) - v_1(\theta_{1k}, \theta_{2\ell})) \\ & \geq \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) (v_1(\theta_{1m}, \theta_{2m}) - v_1(\theta_{1k}, \theta_{2m})) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) (v_2(\theta_{11}, \theta_{2\ell}) - v_2(\theta_{11}, \theta_{21})). \end{aligned} \quad (1)$$

The EAW condition requires the ex ante gains from trade to be sufficient to prevent agents from deviating. The detailed argument is as follows. In order to trade at the highest possible price, the seller could deviate by reporting her highest type θ_{1m} . To prevent this deviation, the designer must compensate her by this amount:

$$\mu(\theta_{1k}, \theta_{2m}) (v_1(\theta_{1m}, \theta_{2m}) - v_1(\theta_{1k}, \theta_{2m})),$$

which is the ex ante value difference between the true type θ_{1k} and the highest type θ_{1m} conditional on the buyer's type being the highest one θ_{2m} . A similar argument can be made for the buyer as well.

If the EAW condition is violated, by contrast, the designer cannot stop the agents from deviating by redistributing the ex ante gains from trade, implying that BIC would be violated. See also Makowski and Mezzetti (1994), Williams (1999), and Krishna and Perry (2000) for similar conditions for the general mechanism design problem with private values and independent beliefs in a continuous type space.

We are ready to present the main result.

Theorem 1. Suppose that there exists a desired mechanism (x^*, t) satisfying NTNP such that the efficient decision rule x^* is described in Table 2. Then, the EAW condition holds.

Proof. In order for the mechanism (x^*, t) to satisfy BB, we let $t_1(\theta_{1k}, \theta_{2\ell}) = -t_2(\theta_{1k}, \theta_{2\ell}) = t(\theta_{1k}, \theta_{2\ell})$ for all $(\theta_{1k}, \theta_{2\ell}) \in \Theta$ throughout the proof, i.e., $t(\theta_{1k}, \theta_{2\ell})$ is the amount

of money the buyer pays the seller when the type report is $(\theta_{1k}, \theta_{2\ell})$. Furthermore, recall that $x^*(\theta_{1k}, \theta_{2\ell}) = 0$ if and only if $k > \ell$. Then, by NTNP, we have $t(\theta_{1k}, \theta_{2\ell}) = 0$ whenever $k > \ell$.

We write down the IIR constraints: $IR_{\theta_{1m}}$ and $IR_{\theta_{21}}$, where $IR_{\theta_{ik}}$ denotes the IIR constraint for agent i of type θ_{ik} :

$$\begin{aligned} IR_{\theta_{1m}} : \quad & \mu(\theta_{2m}|\theta_{1m})t(\theta_{1m}, \theta_{2m}) + \sum_{\ell=1}^{m-1} \mu(\theta_{2\ell}|\theta_{1m})v_1(\theta_{1m}, \theta_{2\ell}) \geq \sum_{\ell=1}^m \mu(\theta_{2\ell}|\theta_{1m})v_1(\theta_{1m}, \theta_{2\ell}); \\ IR_{\theta_{21}} : \quad & \mu(\theta_{11}|\theta_{21})(v_2(\theta_{11}, \theta_{21}) - t(\theta_{11}, \theta_{21})) \geq 0. \end{aligned}$$

The IIR constraints can be simplified to the following:

$$\begin{aligned} IR_{\theta_{1m}} : \quad & t(\theta_{1m}, \theta_{2m}) - v_1(\theta_{1m}, \theta_{2m}) \geq 0; \\ IR_{\theta_{21}} : \quad & v_2(\theta_{11}, \theta_{21}) - t(\theta_{11}, \theta_{21}) \geq 0. \end{aligned}$$

Furthermore, the BIC constraints which stop the seller of type θ_{1k} with $k \neq m$ from deviating to θ_{1m} is as follows: for any $k \neq m$,

$$\begin{aligned} IC_{\theta_{1k} \rightarrow \theta_{1m}} : \quad & \sum_{\ell=k}^m \mu(\theta_{2\ell}|\theta_{1k})t(\theta_{1k}, \theta_{2\ell}) + \sum_{\ell=1}^{k-1} \mu(\theta_{2\ell}|\theta_{1k})v_1(\theta_{1k}, \theta_{2\ell}) \\ & \geq \mu(\theta_{2m}|\theta_{1k})t(\theta_{1m}, \theta_{2m}) + \sum_{\ell=1}^{m-1} \mu(\theta_{2\ell}|\theta_{1k})v_1(\theta_{1k}, \theta_{2\ell}) \\ \Rightarrow \quad & \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell})t(\theta_{1k}, \theta_{2\ell}) - \mu(\theta_{1k}, \theta_{2m})t(\theta_{1m}, \theta_{2m}) \geq \sum_{\ell=k}^{m-1} \mu(\theta_{1k}, \theta_{2\ell})v_1(\theta_{1k}, \theta_{2\ell}). \end{aligned}$$

Similarly, the BIC constraints which stop the buyer of type $\theta_{2\ell}$ with $\ell \neq 1$ from deviating to θ_{21} is as follows: for any $\ell \neq 1$,

$$\begin{aligned} IC_{\theta_{2\ell} \rightarrow \theta_{21}} : \quad & \sum_{k=1}^{\ell} \mu(\theta_{1k}|\theta_{2\ell})(v_2(\theta_{1k}, \theta_{2\ell}) - t(\theta_{1k}, \theta_{2\ell})) \geq \mu(\theta_{11}|\theta_{2\ell})(v_2(\theta_{11}, \theta_{2\ell}) - t(\theta_{11}, \theta_{21})) \\ \Rightarrow \quad & \sum_{k=2}^{\ell} \mu(\theta_{1k}, \theta_{2\ell})v_2(\theta_{1k}, \theta_{2\ell}) - \sum_{k=1}^{\ell} \mu(\theta_{1k}, \theta_{2\ell})t(\theta_{1k}, \theta_{2\ell}) + \mu(\theta_{11}, \theta_{2\ell})t(\theta_{11}, \theta_{21}) \geq 0. \end{aligned}$$

In the rest of the proof, we utilize the above BIC and IIR constraints to verify that the EAW condition must hold.

First, we multiply $IR_{\theta_{1m}}$ by $\sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m})$ to obtain

$$\sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m})(t(\theta_{1m}, \theta_{2m}) - v_1(\theta_{1m}, \theta_{2m})) \geq 0. \quad (2)$$

Second, we multiply $IR_{\theta_{21}}$ by $\sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell})$ to obtain

$$\sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) (v_2(\theta_{11}, \theta_{21}) - t(\theta_{11}, \theta_{21})) \geq 0. \quad (3)$$

Third, summing up $IC_{\theta_{11} \rightarrow \theta_{1m}}, IC_{\theta_{12} \rightarrow \theta_{1m}}, \dots, IC_{\theta_{1,m-1} \rightarrow \theta_{1m}}$, we have

$$\sum_{k=1}^{m-1} \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) - \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) t(\theta_{1m}, \theta_{2m}) - \sum_{k=1}^{m-1} \sum_{\ell=k}^{m-1} \mu(\theta_{1k}, \theta_{2\ell}) v_1(\theta_{1k}, \theta_{2\ell}) \geq 0. \quad (4)$$

Notice that the first term of the left-hand side of inequality (4) can be decomposed into the following:

$$\begin{aligned} & \sum_{k=1}^{m-1} \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) \\ = & \sum_{k=2}^{m-1} \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) t(\theta_{11}, \theta_{2\ell}) + \mu(\theta_{11}, \theta_{21}) t(\theta_{11}, \theta_{21}). \end{aligned}$$

Substituting it back into inequality (4), we obtain

$$\begin{aligned} & \sum_{k=2}^{m-1} \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) t(\theta_{11}, \theta_{2\ell}) + \mu(\theta_{11}, \theta_{21}) t(\theta_{11}, \theta_{21}) \\ & - \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) t(\theta_{1m}, \theta_{2m}) - \sum_{k=1}^{m-1} \sum_{\ell=k}^{m-1} \mu(\theta_{1k}, \theta_{2\ell}) v_1(\theta_{1k}, \theta_{2\ell}) \geq 0. \end{aligned} \quad (5)$$

Fourth, summing up $IC_{\theta_{22} \rightarrow \theta_{21}}, IC_{\theta_{23} \rightarrow \theta_{21}}, \dots, IC_{\theta_{2m} \rightarrow \theta_{21}}$, we have

$$\sum_{\ell=2}^m \sum_{k=2}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) v_2(\theta_{1k}, \theta_{2\ell}) - \sum_{\ell=2}^m \sum_{k=1}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) t(\theta_{11}, \theta_{21}) \geq 0. \quad (6)$$

Notice that the second term of the left-hand side in inequality (6) can be decomposed into the following:

$$\begin{aligned} & - \sum_{\ell=2}^m \sum_{k=1}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) \\ = & - \sum_{\ell=2}^{m-1} \sum_{k=1}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) - \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) t(\theta_{1k}, \theta_{2m}) - \mu(\theta_{1m}, \theta_{2m}) t(\theta_{1m}, \theta_{2m}). \end{aligned}$$

Substituting it back into inequality (6), we obtain

$$\begin{aligned} & \sum_{\ell=2}^m \sum_{k=2}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) v_2(\theta_{1k}, \theta_{2\ell}) - \sum_{\ell=2}^{m-1} \sum_{k=1}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) t(\theta_{1k}, \theta_{2\ell}) - \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) t(\theta_{1k}, \theta_{2m}) \\ & - \mu(\theta_{1m}, \theta_{2m}) t(\theta_{1m}, \theta_{2m}) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) t(\theta_{11}, \theta_{21}) \geq 0. \end{aligned} \quad (7)$$

Fifth, we add up inequalities (2), (3), (5) and (7). After adding up these inequalities, we show that all terms involving the transfers $\{t(\theta_{1k}, \theta_{2\ell})\}_{(k,\ell):k \leq \ell}$ are cancelled out. To see this, we divide the payment into three groups: $t(\theta_{11}, \theta_{21})$, $t(\theta_{1m}, \theta_{2m})$, and the transfers at all the other type profiles.

Observe that $t(\theta_{11}, \theta_{21})$ appears in (i) inequality (3) with its coefficient $-\sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell})$; (ii) inequality (5) with its coefficient $\mu(\theta_{11}, \theta_{21})$; and (iii) inequality (7) with its coefficient $\sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell})$. The summation of the coefficients equals

$$-\sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) + \mu(\theta_{11}, \theta_{21}) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) = 0.$$

Hence, the terms involving $t(\theta_{11}, \theta_{21})$ are cancelled out.

Observe that $t(\theta_{1m}, \theta_{2m})$ appears in (i) inequality (2) with its coefficient $\sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m})$; (ii) inequality (5) with its coefficient $-\sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m})$; and (iii) inequality (7) with its coefficient $-\mu(\theta_{1m}, \theta_{2m})$. The summation of the coefficients equals

$$\sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m}) - \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) - \mu(\theta_{1m}, \theta_{2m}) = 0.$$

Hence, the terms involving $t(\theta_{1m}, \theta_{2m})$ are cancelled out.

Observe that any other arbitrary payment $t(\theta_{1k}, \theta_{2\ell})$ appears in (i) inequality (5) with its coefficient $\mu(\theta_{1k}, \theta_{2\ell})$; and (ii) inequality (7) with its coefficient $-\mu(\theta_{1k}, \theta_{2\ell})$. Since the summation of the coefficients equals

$$\mu(\theta_{1k}, \theta_{2\ell}) - \mu(\theta_{1k}, \theta_{2\ell}) = 0,$$

the terms involving all the other payments $t(\theta_{1k}, \theta_{2\ell})$ are cancelled out.

Therefore, the summation of the left-hand sides of the following inequalities (2), (3), (5) and (7) leads to

$$\begin{aligned} & \sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) v_2(\theta_{11}, \theta_{21}) + \sum_{\ell=2}^m \sum_{k=2}^{\ell} \mu(\theta_{1k}, \theta_{2\ell}) v_2(\theta_{1k}, \theta_{2\ell}) \\ & \geq \sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m}) v_1(\theta_{1m}, \theta_{2m}) + \sum_{k=1}^{m-1} \sum_{\ell=k}^{m-1} \mu(\theta_{1k}, \theta_{2\ell}) v_1(\theta_{1k}, \theta_{2\ell}), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} & \sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) v_2(\theta_{11}, \theta_{21}) + \sum_{k=2}^m \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) v_2(\theta_{1k}, \theta_{2\ell}) \\ & \geq \sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m}) v_1(\theta_{1m}, \theta_{2m}) + \sum_{k=1}^{m-1} \sum_{\ell=k}^{m-1} \mu(\theta_{1k}, \theta_{2\ell}) v_1(\theta_{1k}, \theta_{2\ell}). \end{aligned}$$

After rearrangement, it can be further rewritten as

$$\begin{aligned}
& \sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) v_2(\theta_{11}, \theta_{21}) + \sum_{k=1}^m \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) v_2(\theta_{1k}, \theta_{2\ell}) - \sum_{\ell=1}^m \mu(\theta_{11}, \theta_{2\ell}) v_2(\theta_{11}, \theta_{2\ell}) \\
& \geq \sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m}) v_1(\theta_{1m}, \theta_{2m}) + \sum_{k=1}^m \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) v_1(\theta_{1k}, \theta_{2\ell}) - \sum_{k=1}^m \mu(\theta_{1k}, \theta_{2m}) v_1(\theta_{1k}, \theta_{2m}) \\
& \Rightarrow \sum_{k=1}^m \sum_{\ell=k}^m \mu(\theta_{1k}, \theta_{2\ell}) (v_2(\theta_{1k}, \theta_{2\ell}) - v_1(\theta_{1k}, \theta_{2\ell})) \\
& \geq \sum_{k=1}^{m-1} \mu(\theta_{1k}, \theta_{2m}) (v_1(\theta_{1m}, \theta_{2m}) - v_1(\theta_{1k}, \theta_{2m})) + \sum_{\ell=2}^m \mu(\theta_{11}, \theta_{2\ell}) (v_2(\theta_{11}, \theta_{2\ell}) - v_2(\theta_{11}, \theta_{21})),
\end{aligned}$$

which is identical to the EAW condition. This completes the proof of the theorem. $\square$

5 The Implications in a Continuous Type Space

This section analyzes bilateral trade with a continuous type space through the lens of the ex ante welfare (EAW) condition.

5.1 Preliminaries

We consider $T^{Con} = ([0, 1]^2, f)$ as the *benchmark continuous type space* where each agent's type space is $[0, 1]$ and f is a full-support continuous density function over $[0, 1]^2$.

We start with the space of all bounded, real-valued functions on $[0, 1]^2$, denoted as $B([0, 1]^2)$. This space is equipped with the supremum norm:

$$\|v\|_{\infty} = \sup_{\theta \in [0, 1]^2} |v(\theta)|.$$

By the Hewitt-Yosida theorem, the continuous dual of $B([0, 1]^2)$ is the space of all bounded finitely additive measures on the field generated by the subsets of $[0, 1]^2$. We denote this space by $ba([0, 1]^2)$.

A sequence of measures μ_n converges to μ in the weak*-topology if, for all $v \in B([0, 1]^2)$,

$$\int_{\theta \in [0, 1]^2} v(\theta) d\mu_n(\theta) \rightarrow \int_{\theta \in [0, 1]^2} v(\theta) d\mu(\theta).$$

Let $\mathcal{P} \subset ba([0, 1]^2)$ be the space of probability measures such that for any $\mu \in ba([0, 1]^2)$, the following two properties hold: (i) nonnegativity: $\mu(A) \geq 0$ for any

A in the field and (ii) normalization: $\mu([0, 1]^2) = 1$. We then endow $\mathcal{P}$ with the topology induced by the convergence of integrals in the weak*-topology.

We say that $\Theta^{(m)} \subseteq [0, 1]^2$ is the m -finite grid types such that for each $\theta = (\theta_1, \theta_2) \in \Theta^{(m)}$ and $i = \{1, 2\}$, there exists $k \in \{1, \dots, m\}$ such that

$$\theta_{ik} = \frac{k-1}{m-1}.$$

That is, both agents have the same number of types, denoted by m . The types are equally spaced on the interval $[0, 1]$. For each $m \in \mathbb{N}$, we denote by $T^{(m)} = (\Theta^{(m)}, \mu^{(m)})$ the discretized type space, where $\mu^{(m)}$ is a probability mass function on $\Theta^{(m)}$. We also denote by $\mathcal{P}(\Theta^{(m)})$ the simplex of probability mass functions on $\Theta^{(m)}$. We introduce a metric between two probability mass functions: for all $\mu_1^{(m)}, \mu_2^{(m)} \in \mathcal{P}(\Theta^{(m)})$,

$$d(\mu_1^{(m)}, \mu_2^{(m)}) = \frac{1}{2} \sum_{k=1}^m \sum_{\ell=1}^m \left| \mu_1^{(m)}(\theta_{1k}, \theta_{2\ell}) - \mu_2^{(m)}(\theta_{1k}, \theta_{2\ell}) \right|.$$

We further define the following product space:

$$\mathcal{P}^\infty \equiv \prod_{m=1}^{\infty} \mathcal{P}(\Theta^{(m)}).$$

Then, we say that a sequence of probability mass functions $\{\mu^{(m)}\}_{m \in \mathbb{N}} \in \mathcal{P}^\infty$ converges to $f \in \mathcal{P}$ if, as $m \rightarrow \infty$,

$$d(\mu^{(m)}, f^{(m)}) \rightarrow 0,$$

where $f^{(m)}$ denotes the probability mass function that truncates f to $\Theta^{(m)}$ as follows:

$$f^{(m)}(\theta_{1k}, \theta_{2\ell}) = \begin{cases} \int_{[\theta_{1,k-1}, \theta_{1k}] \times [\theta_{2,\ell-1}, \theta_{2\ell}]} f(\theta) d\theta & \text{if } (k, \ell) \in \{2, \dots, m\}^2 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify

$$\sum_{k=1}^m \sum_{\ell=1}^m f^{(m)}(\theta_{1k}, \theta_{2\ell}) = \sum_{k=2}^m \sum_{\ell=2}^m \int_{[\theta_{1,k-1}, \theta_{1k}] \times [\theta_{2,\ell-1}, \theta_{2\ell}]} f(\theta) d\theta = 1.$$

This class of convergent sequences introduces a topology on $\mathcal{P}^\infty$ that allows us to approximate any sequence of probability measures in the space $\mathcal{P}$ in terms of a sequence of probability mass functions in $\mathcal{P}^\infty$.

5.2 The Limit Ex Ante Welfare Condition

We now propose the *limit ex ante welfare* (limit EAW) condition as its extended condition to a continuous type space.

Definition 9. The collection of valuations $(v_1(\theta), v_2(\theta))_{\theta \in [0,1]^2}$ and the benchmark continuous type space $T^{Con} = ([0,1]^2, f)$ satisfy *the limit ex ante welfare (limit EAW) condition* if the following inequality holds:

$$\begin{aligned} & \int_0^1 \int_0^{\theta_2} (v_2(\theta_1, \theta_2) - v_1(\theta_1, \theta_2)) f(\theta_1, \theta_2) d\theta_1 d\theta_2 \\ & \geq \int_0^1 (v_1(1, 1) - v_1(\theta_1, 1)) f(\theta_1, 1) d\theta_1 + \int_0^1 (v_2(0, \theta_2) - v_2(0, 0)) f(0, \theta_2) d\theta_2. \end{aligned}$$

We show that the limit EAW condition holds trivially in the benchmark continuous type space.

Lemma 1. For any collection of valuations $(v_1(\theta), v_2(\theta))_{\theta \in [0,1]^2}$ and any benchmark continuous type space $T^{Con} = ([0,1]^2, f)$, the limit EAW condition is always satisfied.

Proof. Since $v_1(\cdot)$ is strictly increasing in θ_1 , we have $v_1(1, 1) - v_1(\theta_1, 1) \geq 0$ for all $\theta_1 \in [0, 1]$, implying

$$\int_0^1 (v_1(1, 1) - v_1(\theta_1, 1)) f(\theta_1, 1) d\theta_1 \geq 0.$$

On the other hand, we also have

$$\int_0^1 (v_1(1, 1) - v_1(\theta_1, 1)) f(\theta_1, 1) d\theta_1 \leq (v_1(1, 1) - v_1(0, 1)) \int_0^1 f(\theta_1, 1) d\theta_1 = 0,$$

where the weak inequality holds because $v_1(\theta_1, 1) \geq v_1(0, 1)$ for all $\theta_1 \in [0, 1]$ and the equality holds because $\int_0^1 f(\theta_1, 1) d\theta_1 = 0$. Therefore, we obtain

$$\int_0^1 (v_1(1, 1) - v_1(\theta_1, 1)) f(\theta_1, 1) d\theta_1 = 0.$$

Similarly, since $v_2(\cdot)$ is strictly increasing in θ_2 , we have $v_2(0, \theta_2) - v_2(0, 0) \geq 0$ for all $\theta_2 \in [0, 1]$, implying

$$\int_0^1 (v_2(0, \theta_2) - v_2(0, 0)) f(0, \theta_2) d\theta_2 \geq 0.$$

On the other hand, we also have

$$\int_0^1 (v_2(0, \theta_2) - v_2(0, 0)) f(0, \theta_2) d\theta_2 \leq (v_2(0, 1) - v_2(0, 0)) \int_0^1 f(0, \theta_2) d\theta_2 = 0,$$

where the weak inequality holds because $v_2(0, \theta_2) \leq v_2(0, 1)$ for all $\theta_2 \in [0, 1]$ and the equality holds because $\int_0^1 f(0, \theta_2) d\theta_2 = 0$. Therefore, we obtain

$$\int_0^1 (v_2(0, \theta_2) - v_2(0, 0)) f(0, \theta_2) d\theta_2 = 0.$$

We conclude that the right-hand side of the inequality in the limit EAW condition is zero. In contrast, its left-hand side represents the ex ante gains from trade, which are always nonnegative. Therefore, the limit EAW condition is always satisfied. This completes the proof. $\square$

5.3 Locally Robust Implementation

In this subsection, we further assume that the agents' valuation functions are linear, which are given by, for each type profile $(\theta_{1k}, \theta_{2\ell}) \in [0, 1]^2$,

$$v_1(\theta_{1k}, \theta_{2\ell}) = \theta_{1k} + \gamma\theta_{2\ell} \text{ and } v_2(\theta_{1k}, \theta_{2\ell}) = \theta_{2\ell} + \gamma\theta_{1k},$$

where $\gamma \in [0, 1)$ denotes the common degree of interdependence. Since both agents have the same degree of interdependence, we have

$$v_2(\theta_{1k}, \theta_{2\ell}) - v_1(\theta_{1k}, \theta_{2\ell}) = (1 - \gamma)(\theta_{2\ell} - \theta_{1k}),$$

implying that $v_2(\theta_{1k}, \theta_{2\ell}) \geq v_1(\theta_{1k}, \theta_{2\ell})$ if and only if $\theta_{2\ell} \geq \theta_{1k}$, or equivalently, $\ell \geq k$. Therefore, the efficient decision rule in this case is identical to the one in Table 2, and thus, the EAW condition is necessary for the existence of desired mechanisms satisfying NTNP.

Let

$$\mathcal{P}_\varepsilon^\infty = \{ \{ \mu^{(m)} \}_{m \in \mathbb{N}} \in \mathcal{P}^\infty : d(\mu^{(m)}, f^{(m)}) \leq \varepsilon \text{ for all } m \in \mathbb{N} \text{ large enough} \},$$

where $f^{(m)}$ is the probability mass function that truncates the continuous density function f to $\Theta^{(m)}$, as introduced in Subsection 5.1. Let $\Gamma^{(m)} = (\Theta^{(m)}, x^{(m)}, t^{(m)})$ denote a direct mechanism in which $x^{(m)} : \Theta^{(m)} \rightarrow [0, 1]$ and $t^{(m)} : \Theta^{(m)} \rightarrow \mathbb{R}^2$ are the decision rule and transfer rule, respectively.

Definition 10. Efficient bilateral trade is ε -robust implementable for the benchmark continuous type space $T^{Con} = ([0, 1]^2, f)$ if, for any $\{ \mu^{(m)} \}_{m \in \mathbb{N}} \in \mathcal{P}_\varepsilon^\infty$, there exists a sequence of mechanisms $\{ \Gamma^{(m)} \}_{m=1}^\infty$ such that, for each $m \in \mathbb{N}$, the mechanism $\Gamma^{(m)} = (\Theta^{(m)}, x^{(m)}, t^{(m)})$ satisfies BIC, IIR, EFF, BB, and NTNP on the discrete type space $T^{(m)} = (\Theta^{(m)}, \mu^{(m)})$.

Our next main result presents the impossibility of ε -robust implementability. To achieve this, we employ the EAW condition, which serves as a necessary condition for the existence of desired mechanisms that satisfy NTNP. The key is to show that the EAW condition is violated.

We associate each finite economy indexed by m with $(\gamma, \Theta^{(m)}, \mu^{(m)})$, where $v_1(\theta_{1k}, \theta_{2\ell}) = \theta_{1k} + \gamma\theta_{2\ell}$, $v_2(\theta_{1k}, \theta_{2\ell}) = \theta_{2\ell} + \gamma\theta_{1k}$, and $T^{(m)} = (\Theta^{(m)}, \mu^{(m)})$ is a finite type space that approximates the benchmark type space $T^{Con} = ([0, 1]^2, f)$.

Theorem 2 (Impossibility of Locally Robust Implementation). Suppose that the agents' valuation functions are given by $v_1(\theta_{1k}, \theta_{2\ell}) = \theta_{1k} + \gamma\theta_{2\ell}$ and $v_2(\theta_{1k}, \theta_{2\ell}) = \theta_{2\ell} + \gamma\theta_{1k}$, where $\gamma \in [0, 1)$. Then, for each $\varepsilon > 0$, there exists $\bar{\gamma}_\varepsilon \in (0, 1)$ such that efficient bilateral trade associated with any $\gamma \in (\bar{\gamma}_\varepsilon, 1)$ fails to be ε -robust implementable for the benchmark continuous type space $T^{Con} = ([0, 1]^2, f)$.

Proof. The proof is in the Appendix. $\square$

This result shows that, if the degree of interdependence is high enough, we can construct a perturbation $\{\mu^{(m)}\}_{m \in \mathbb{N}}$ around the benchmark continuous type space $T^{Con} = ([0, 1]^2, f)$ such that the EAW condition fails for a nearby finite environment.

The idea behind Theorem 2 is analogous to Lopomo et al. (2022) who study a robust version of the classic surplus extraction problem. In their model, the designer knows only that the beliefs of each type of the agents belong to some set, and design mechanisms that are suitable for all possible beliefs in that set.⁷ As demonstrated in the proof of Theorem 2, the uncertainty underlying locally robust implementation is captured by a set of beliefs formed by ε -contamination of the truncated beliefs of f , $\{f^{(m)}\}_{m \in \mathbb{N}}$, that is,

$$\hat{\mu}_\varepsilon^{(m)} = (1 - \varepsilon)f^{(m)} + \varepsilon\mu_S^{(m)},$$

where $\mu_S^{(m)}$ is some singular distribution, which is explicitly specified in the proof of Theorem 2. We assume that the designer cannot distinguish between $\{f^{(m)}\}_{m \in \mathbb{N}}$ and $\{\hat{\mu}_\varepsilon^{(m)}\}_{m \in \mathbb{N}}$, as long as they are sufficiently close to each other topologically. We then show that the possibility result for the existence of desired mechanisms satisfying NTNP is not robust to this ε -contamination.

⁷As noted in Lopomo et al. (2022), the behavioral foundation of our analysis traces back to Knightian decision theory developed in Bewley (1986). The reader is referred to Bewley (1986) and Lopomo et al. (2022) for detailed discussion of Knightian uncertainty.

6 Fieseler, Kittsteiner, and Moldovanu (2003)

In this section, we revisit Theorem 5 of Fieseler, Kittsteiner, and Moldovanu (henceforth, FKM, 2003) and, as we do in Section 5.3, clarify the implication of their Theorem 5 in the specification of $T^{Con} = ([0, 1]^2, f)$, the benchmark continuous type space where each agent's type space is $[0, 1]$ and f is a full-support continuous density function over $[0, 1]^2$. As in Section 5.3, suppose that the agents' valuation functions are linear, which are given by, for each type profile $(\theta_{1k}, \theta_{2\ell}) \in [0, 1]^2$,

$$v_1(\theta_{1k}, \theta_{2\ell}) = \theta_{1k} + \gamma\theta_{2\ell} \text{ and } v_2(\theta_{1k}, \theta_{2\ell}) = \theta_{2\ell} + \gamma\theta_{1k},$$

where $\gamma \in [0, 1)$ is the common parameter for preference interdependence. For each agent i , let $j \neq i$ and we define

$$f_i(\theta_i) = \int_{\underline{\theta}}^{\bar{\theta}} f(\theta_i, \tilde{\theta}_j) d\tilde{\theta}_j,$$

$$f_{j|i}(\theta_j|\theta_i) = \frac{f(\theta_i, \theta_j)}{f_i(\theta_i)} = \frac{f(\theta_i, \theta_j)}{\int_{\underline{\theta}}^{\bar{\theta}} f(\theta_i, \tilde{\theta}_j) d\tilde{\theta}_j},$$

and

$$F_{j|i}(\theta_j|\theta_i) = \int_{\underline{\theta}}^{\theta_j} f_{j|i}(\tilde{\theta}_j|\theta_i) d\tilde{\theta}_j.$$

To have a good benchmark based on continuous type space, we state Theorem 5 of FKM (2003). It is important to note that Gresik (1991b) already derived a different condition for the existence of desired mechanisms (Theorem 3 of Gresik (1991b)).

Lemma 2 (Theorem 5 in FKM (2003)). Suppose that the types are independently distributed, i.e., for any $i, j \in \{1, 2\}$ with $i \neq j$, $\theta_j \in \Theta_j$, and $\theta_i, \theta'_i \in \Theta_i$, $f_{j|i}(\theta_j|\theta_i) = f_{j|i}(\theta_j|\theta'_i)$. There exists a desired mechanism if and only if either one of the following two conditions hold:

1. there is a price p such that $\int_{\underline{\theta}}^{\bar{\theta}} v_2(\theta_1, \underline{\theta}) f_{1|2}(\theta_1|\underline{\theta}) d\theta_1 \geq p \geq \int_{\underline{\theta}}^{\bar{\theta}} v_1(\bar{\theta}, \theta_2) f_{2|1}(\theta_2|\bar{\theta}) d\theta_2$;
2. $v_2(\theta) \leq v_1(\theta)$ for all $\theta \in \Theta$.⁸

⁸To be precise, FKM (2003) only require ex ante budget surplus rather than ex post budget balance, which we assume. The difference between these two kinds of budget balances disappears in FKM (2003) because Börgers (2015) in Proposition 3.6 and Börgers and Norman (2009) in Proposition 2 show that in the case of independent beliefs, ex ante budget surplus implies ex post budget balance.

In their stylized example, Kunimoto and Zhang (2022) demonstrate that the first condition in Lemma 2 amounts to $\gamma - \gamma \geq 2$, which is impossible. The second condition in Lemma 2 is also impossible because $v_2(\theta_1, \theta_2) \geq v_1(\theta_1, \theta_2)$ if and only if $\theta_2 \geq \theta_1$. Therefore, within this continuous type space setup, which is a prominent one, we conclude that there are no desired mechanisms. Imposing NTNP constraint on desired mechanisms and requiring efficient bilateral trade to be locally robust implementable, we show by our Theorem 2 that, if the degree of interdependent values is sufficiently high, FKM’s (2003) impossibility result in continuous type space can be extended to a impossibility result in settings with correlated beliefs.

7 Conclusion

This paper has characterized when efficient, voluntary bilateral trade can be achieved in environments with interdependent values and correlated beliefs. We model the bilateral trade model as a setting where each agent’s private information can affect both valuations, and where their beliefs about the opponent’s types may be correlated.

Our analysis identified the ex ante welfare (EAW) condition as a necessary requirement to implement ex post efficient bilateral trade in any finite type space. Leveraging this condition, we investigated the implications of efficient trade in continuous type spaces. By taking the limit of the EAW condition as the finite type space approximates the benchmark continuous type space, we derived a corresponding limit EAW condition.

We demonstrated that while this limit condition holds trivially in benchmark continuous settings with a full-support density function (Lemma 1), its validity is challenged when subjected to a robustness check. Specifically, we introduced the notion of locally robust implementation, which requires efficient trade to be implementable uniformly across all finite type spaces that approximate the benchmark continuous type space. Our Theorem 2 established that under high interdependence, the EAW condition fails in a finite type space that approximate the benchmark continuous type space; that is, locally robust implementation of efficient trade is impossible.

These findings provide a new perspective on the fundamental impossibility results established by Myerson and Satterthwaite (1983) for private values and Fieseler, Kittsteiner, and Moldovanu (FKM, 2003) for interdependent values, but both under independent beliefs. We show that these classic negative results are even more far-reaching than previously understood: when the degree of interde-

pendent values is sufficiently high, impossibility emerges as the robust limit of discrete environments, even under belief correlation. Extending these insights to broader mechanism design settings remains an important open question. We hope our characterization serves as a meaningful step toward a complete understanding of robust implications of efficient mechanisms.

8 Appendix

In the Appendix, we provide all the proofs omitted from the main text of the paper.

8.1 Proof of Proposition 1

Proof. The proof consists of four steps. In Step 1, we show that $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$. In Step 2, we show that $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} . Finally, Steps 3 and 4 pin down the unique efficient decision rule $x^*(\cdot)$.

Step 1: $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ for every $k \in \{1, \dots, m\}$.

Proof. Fix an arbitrary $\theta_{1k} \in \Theta_1$. Suppose, on the contrary, that there exist $\theta_{2\ell}, \theta_{2\ell'} \in \Theta_2$ with $\ell < \ell'$ such that $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ and $x^*(\theta_{1k}, \theta_{2\ell'}) = 0$. Notice that $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ implies $v_2(\theta_{1k}, \theta_{2\ell}) \geq v_1(\theta_{1k}, \theta_{2\ell})$. If the buyer's type increases to $\theta_{2\ell'}$, by the single crossing condition, we have

$$\begin{aligned} v_2(\theta_{1k}, \theta_{2\ell'}) - v_2(\theta_{1k}, \theta_{2\ell}) &> v_1(\theta_{1k}, \theta_{2\ell'}) - v_1(\theta_{1k}, \theta_{2\ell}) \\ \Rightarrow v_2(\theta_{1k}, \theta_{2\ell'}) - v_1(\theta_{1k}, \theta_{2\ell'}) &> v_2(\theta_{1k}, \theta_{2\ell}) - v_1(\theta_{1k}, \theta_{2\ell}), \end{aligned}$$

implying $v_2(\theta_{1k}, \theta_{2\ell'}) > v_1(\theta_{1k}, \theta_{2\ell'})$ must hold. Therefore, we obtain $x^*(\theta_{1k}, \theta_{2\ell'}) = 1$, contradicting our hypothesis that $x^*(\theta_{1k}, \theta_{2\ell'}) = 0$. This completes the proof of Step 1. $\square$

Step 2: $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} for every $\ell \in \{1, \dots, m\}$.

Proof. Fix an arbitrary $\theta_{2\ell} \in \Theta_2$. Suppose, on the contrary, that there exist $\theta_{1k}, \theta_{1k'} \in \Theta_1$ with $k < k'$ such that $x^*(\theta_{1k}, \theta_{2\ell}) = 0$ and $x^*(\theta_{1k'}, \theta_{2\ell}) = 1$. Notice that $x^*(\theta_{1k}, \theta_{2\ell}) = 0$ implies $v_1(\theta_{1k}, \theta_{2\ell}) > v_2(\theta_{1k}, \theta_{2\ell})$. If the seller's type increases to $\theta_{1k'}$, by the single crossing condition, we have

$$\begin{aligned} v_1(\theta_{1k'}, \theta_{2\ell}) - v_1(\theta_{1k}, \theta_{2\ell}) &> v_2(\theta_{1k'}, \theta_{2\ell}) - v_2(\theta_{1k}, \theta_{2\ell}) \\ \Rightarrow v_1(\theta_{1k'}, \theta_{2\ell}) - v_2(\theta_{1k'}, \theta_{2\ell}) &> v_1(\theta_{1k}, \theta_{2\ell}) - v_2(\theta_{1k}, \theta_{2\ell}), \end{aligned}$$

implying $v_1(\theta_{1k'}, \theta_{2\ell}) > v_2(\theta_{1k'}, \theta_{2\ell})$ must hold. Therefore, we obtain $x^*(\theta_{1k'}, \theta_{2\ell}) = 0$, contradicting our hypothesis $x^*(\theta_{1k'}, \theta_{2\ell}) = 1$. This completes the proof of Step 2. $\square$

Step 3: For all $k \in \{2, \dots, m\}$,

$$x^*(\theta_{1k}, \theta_{2k}) = 1 \quad (8)$$

and

$$x^*(\theta_{1k}, \theta_{2,k-1}) = 0. \quad (9)$$

Proof. We highlight that the only place where responsiveness plays a role lies in Step 3. We prove this by induction.

First, we show that equations (8) and (9) hold when $k = 2$. Suppose, on the contrary, that $x^*(\theta_{12}, \theta_{21}) = 1$. Since $x^*(\theta_{11}, \theta_{21}) = x^*(\theta_{12}, \theta_{21}) = 1$ and $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ for each θ_{1k} (by Step 1), we have $x^*(\theta_{11}, \theta_{2\ell}) = x^*(\theta_{12}, \theta_{2\ell}) = 1$ for all $\ell \in \{1, \dots, m\}$, violating the responsiveness of the efficient decision rule. Hence, $x^*(\theta_{12}, \theta_{21}) = 0$ must hold.

Moreover, suppose, on the contrary, that $x^*(\theta_{12}, \theta_{22}) = 0$. Since $x^*(\theta_{11}, \theta_{21}) = 1$ and $x^*(\theta_{11}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$, we have $x^*(\theta_{11}, \theta_{22}) = x^*(\theta_{11}, \theta_{21}) = 1$. Moreover, since $x^*(\theta_{12}, \theta_{21}) = x^*(\theta_{12}, \theta_{22}) = 0$ and $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} for each $\theta_{2\ell}$, we have $x^*(\theta_{1k}, \theta_{21}) = x^*(\theta_{1k}, \theta_{22}) = 0$ for all $k \geq 2$. To summarize, we have

$$x^*(\theta_{1k}, \theta_{21}) = x^*(\theta_{1k}, \theta_{22}) \text{ for all } k \in \{1, \dots, m\},$$

which violates the responsiveness of the efficient decision rule x^* . Hence, $x^*(\theta_{12}, \theta_{22}) = 1$ must hold.

Fix $\tilde{k} \in \{2, \dots, m\}$ arbitrarily. By our induction hypothesis, we assume that $x^*(\theta_{1k}, \theta_{2k}) = 1$ and $x^*(\theta_{1k}, \theta_{2,k-1}) = 0$ for each $k \in \{2, \dots, \tilde{k}\}$. We will show that both equations still hold when $k = \tilde{k} + 1$. Suppose, on the contrary, that $x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}}) = 1$. Since $x^*(\theta_{1,\tilde{k}}, \theta_{2,\tilde{k}-1}) = 0$ (induction hypothesis) and $x^*(\theta_{1k}, \theta_{2,\tilde{k}-1})$ is non-increasing in θ_{1k} (by Step 2), we have $x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}-1}) = x^*(\theta_{1,\tilde{k}}, \theta_{2,\tilde{k}-1}) = 0$. Furthermore, since $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ for each θ_{1k} (by Step 1), we obtain

$$x^*(\theta_{1,\tilde{k}}, \theta_{2\ell}) = x^*(\theta_{1,\tilde{k}+1}, \theta_{2\ell}) = 0$$

for all $\ell \leq \tilde{k} - 1$. Finally, since $x^*(\theta_{1,\tilde{k}}, \theta_{2,\tilde{k}}) = x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}}) = 1$ and $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ for each θ_{1k} (by Step 1), we have

$$x^*(\theta_{1,\tilde{k}}, \theta_{2\ell}) = x^*(\theta_{1,\tilde{k}+1}, \theta_{2\ell}) = 1$$

for all $\ell \geq \tilde{k}$. To summarize, we obtain

$$x^*(\theta_{1\tilde{k}}, \theta_{2\ell}) = x^*(\theta_{1,\tilde{k}+1}, \theta_{2\ell}) \text{ for all } \ell \in \{1, \dots, m\},$$

which violates the responsiveness of the efficient decision rule x^* . Hence, $x^*(\theta_{1,\tilde{k}+1}, \theta_{2\tilde{k}}) = 0$ must hold.

Lastly, suppose, on the contrary, that $x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}+1}) = 0$. Since $x^*(\theta_{1\tilde{k}}, \theta_{2\tilde{k}}) = 1$ (induction hypothesis) and $x^*(\theta_{1\tilde{k}}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ (by Step 1), we have $x^*(\theta_{1\tilde{k}}, \theta_{2,\tilde{k}+1}) = x^*(\theta_{1\tilde{k}}, \theta_{2\tilde{k}}) = 1$. Moreover, since $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} for each $\theta_{2\ell}$ (by Step 2), we obtain

$$x^*(\theta_{1k}, \theta_{2\tilde{k}}) = x^*(\theta_{1k}, \theta_{2,\tilde{k}+1}) = 1$$

for all $k \leq \tilde{k}$. Finally, since $x^*(\theta_{1,\tilde{k}+1}, \theta_{2\tilde{k}}) = x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}+1}) = 0$ and $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} for each $\theta_{2\ell}$ (by Step 2), we have

$$x^*(\theta_{1k}, \theta_{2\tilde{k}}) = x^*(\theta_{1k}, \theta_{2,\tilde{k}+1}) = 0$$

for all $k \geq \tilde{k} + 1$. To summarize, we obtain

$$x^*(\theta_{1k}, \theta_{2\tilde{k}}) = x^*(\theta_{1k}, \theta_{2,\tilde{k}+1}) \text{ for all } k \in \{1, \dots, m\},$$

which violates the responsiveness of the efficient decision rule x^* . Hence, $x^*(\theta_{1,\tilde{k}+1}, \theta_{2,\tilde{k}+1}) = 1$ must hold.

We conclude that $x^*(\theta_{1k}, \theta_{2k}) = 1$ and $x^*(\theta_{1k}, \theta_{2,k-1}) = 0$ are satisfied for all $k \in \{2, \dots, m\}$. $\square$

Step 4: For each $(\theta_{1k}, \theta_{2\ell}) \in \Theta$, $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ if and only if $k \leq \ell$.

Proof. The following table describes the allocation decisions implied by Step 3.

Table 3

$x^*(\cdot)$	θ_{21}	θ_{22}	θ_{23}	$\cdots$	$\theta_{2,m-1}$	θ_{2m}
θ_{11}						
θ_{12}	0	1				
θ_{13}		0	1			
$\vdots$			$\ddots$	$\ddots$		
$\theta_{1,m-1}$				$\ddots$	$\ddots$	
θ_{1m}					0	1

For all $k \in \{2, \dots, m\}$, since $x^*(\theta_{1k}, \theta_{2k}) = 1$ (by Step 3) and $x^*(\theta_{1k}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ (by Step 1), it follows that $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ for all $\ell \geq k$.

Moreover, for all $\ell \in \{1, \dots, m-1\}$, since $x^*(\theta_{1,\ell+1}, \theta_{2\ell}) = 0$ (by Step 3) and $x^*(\theta_{1k}, \theta_{2\ell})$ is non-increasing in θ_{1k} (by Step 2), it follows that $x^*(\theta_{1k}, \theta_{2\ell}) = 0$ for all $k \geq \ell + 1$. Finally, since $x^*(\theta_{11}, \theta_{21}) = 1$ (by our assumption) and $x^*(\theta_{11}, \theta_{2\ell})$ is nondecreasing in $\theta_{2\ell}$ (by Step 1), we obtain $x^*(\theta_{11}, \theta_{2\ell}) = 1$ for all $\ell \geq 1$. To combine all the implications, we conclude that $x^*(\theta_{1k}, \theta_{2\ell}) = 1$ if and only if $k \leq \ell$. $\square$

We thus complete the proof of Proposition 1. $\square$

8.2 Proof of Theorem 2

Proof. For each $m \in \mathbb{N}$ with $m \geq 2$, we define $\mu_S^{(m)}$ to be the following singular distribution:

$$\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) = \begin{cases} \frac{1}{m-1} & \text{if } (k, \ell) = (1, m) \\ \frac{1}{2(m-1)} & \text{if } k = 1 \text{ and } \ell \in \{2, \dots, m-1\} \\ \frac{1}{2(m-1)} & \text{if } \ell = m \text{ and } k \in \{2, \dots, m-1\} \\ 0 & \text{otherwise.} \end{cases}$$

For each $\varepsilon \in (0, 1)$, we then construct the perturbation $\{\widehat{\mu}_\varepsilon^{(m)}\}_{m \in \mathbb{N}} \in \mathcal{P}_\varepsilon^\infty$ with the following property:

$$\widehat{\mu}_\varepsilon^{(m)}(\theta_{1k}, \theta_{2\ell}) = (1 - \varepsilon)f^{(m)}(\theta_{1k}, \theta_{2\ell}) + \varepsilon\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}).$$

The proof is divided into two steps.

Step 1: For every $\varepsilon > 0$,

$$\{\widehat{\mu}_\varepsilon^{(m)}\}_{m \in \mathbb{N}} \in \mathcal{P}_\varepsilon^\infty,$$

or equivalently, there exist $M \in \mathbb{N}$ such that for all $m \geq M$, $d(\widehat{\mu}_\varepsilon^{(m)}, f^{(m)}) \leq \varepsilon$.

Proof. The distance between $\widehat{\mu}_\varepsilon^{(m)}$ and $f^{(m)}$, denoted by $d(\widehat{\mu}_\varepsilon^{(m)}, f^{(m)})$, is computed below:

$$\begin{aligned} d(\widehat{\mu}_\varepsilon^{(m)}, f^{(m)}) &= \frac{1}{2} \sum_{k=1}^m \sum_{\ell=1}^m |\widehat{\mu}_\varepsilon^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell})| \\ &= \frac{1}{2} \sum_{k=1}^m \sum_{\ell=1}^m \left| (1 - \varepsilon)f^{(m)}(\theta_{1k}, \theta_{2\ell}) + \varepsilon\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right| \\ &= \frac{\varepsilon}{2} \sum_{k=1}^m \sum_{\ell=1}^m \left| \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right|. \end{aligned}$$

We partition the set of type profiles $\Theta^{(m)} = \{(\theta_{1k}, \theta_{2\ell}) \mid (k, \ell) \in \{1, \dots, m\}^2\}$ into two subsets: the singular set, denoted by $\Theta_S^{(m)}$, which consists of all boundary points to which the singular distribution $\mu_S^{(m)}$ assigns positive probability mass:

$$\Theta_S^{(m)} = \{(\theta_{1k}, \theta_{2\ell}) \mid (k, \ell) \in \{1\} \times \{2, \dots, m\}\} \cup \{(\theta_{1k}, \theta_{2\ell}) \mid (k, \ell) \in \{2, \dots, m-1\} \times \{m\}\}$$

and the interior set, denoted by $\Theta_I^{(m)}$, which contains all remaining points:

$$\Theta_I^{(m)} = \Theta^{(m)} \setminus \Theta_S^{(m)}.$$

Then, we continue the computation of $d(\hat{\mu}_\varepsilon^{(m)}, f^{(m)})$:

$$\begin{aligned} d(\hat{\mu}_\varepsilon^{(m)}, f^{(m)}) &= \frac{\varepsilon}{2} \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} \left| \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right| \\ &\quad + \frac{\varepsilon}{2} \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_I^{(m)}} \left| \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right| \\ &= \frac{\varepsilon}{2} \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} \left| \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right| \\ &\quad + \frac{\varepsilon}{2} \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_I^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}). \end{aligned} \tag{10}$$

The second equality holds because the singular distribution $\mu_S^{(m)}$ has zero probability mass on the interior points.

We further argue that there exists $\widetilde{M} \in \mathbb{N}$ such that for each $(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}$ and each $m \geq \widetilde{M}$,

$$\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) > 0.$$

Since the density function f integrates to one over $[0, 1]^2$, it must be bounded. Let $f_*^{\max}$ be the maximum value of the density function:

$$f_*^{\max} = \max_{\theta \in [0, 1]^2} f(\theta) < \infty.$$

Then, for each $(k, \ell) \in \{2, \dots, m\}^2$, we obtain

$$f^{(m)}(\theta_{1k}, \theta_{2\ell}) = \int_{[\theta_{1, k-1}, \theta_{1k}] \times [\theta_{2, \ell-1}, \theta_{2\ell}]} f(\theta) d\theta \leq f_*^{\max} \int_{[\theta_{1, k-1}, \theta_{1k}] \times [\theta_{2, \ell-1}, \theta_{2\ell}]} d\theta = \frac{1}{(m-1)^2} f_*^{\max}.$$

As a result, for each $(k, \ell) \in \{2, \dots, m\}^2$, we compute:

$$\begin{aligned} \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) &\geq \frac{1}{2(m-1)} - \frac{1}{(m-1)^2} f_*^{\max} \\ &= \frac{1}{2(m-1)^2} (m-1-2f_*^{\max}). \end{aligned}$$

Since $f_*^{\max} < \infty$, we obtain that, for all $m \in \mathbb{N}$, if $m \geq 1 + 2f_*^{\max}$,

$$\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \geq 0.$$

Similarly, for each $(k, \ell) \in \{1\} \times \{2, \dots, m\}$, we compute:

$$\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) = \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) > 0.$$

To summarize, for all $(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}$ and all $m \geq 1 + 2f_*^{\max}$, we have

$$\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \geq 0.$$

Then, for all $m \geq 1 + 2f_*^{\max}$, expression (10) can be simplified as:

$$\begin{aligned} d(\widehat{\mu}_\varepsilon^{(m)}, f^{(m)}) &= \frac{\varepsilon}{2} \left[\sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} \left(\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) - f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right) + \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_I^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right] \\ &= \frac{\varepsilon}{2} \left[1 - \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) + \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_I^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right] \\ &= \varepsilon \left(1 - \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) \right), \end{aligned}$$

where the second equality follows because $\sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} \mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) = 1$ and the third equality follows because $\sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_I^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) = 1 - \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell})$. Furthermore, notice that

$$\begin{aligned} \sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) &= \sum_{(\theta_{1k}, \theta_{2\ell}) \in \{1\} \times \{2, \dots, m\}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) + \sum_{(\theta_{1k}, \theta_{2\ell}) \in \{2, \dots, m-1\} \times \{m\}} f^{(m)}(\theta_{1k}, \theta_{2\ell}) \\ &= \sum_{(\theta_{1k}, \theta_{2\ell}) \in \{2, \dots, m-1\} \times \{m\}} f^{(m)}(\theta_{1k}, \theta_{2\ell}), \end{aligned}$$

where the second equality follows because $f^{(m)}(\theta_{1k}, \theta_{2\ell}) = 0$ for all $(\theta_{1k}, \theta_{2\ell}) \in \{1\} \times \{2, \dots, m\}$. Since $\sum_{(\theta_{1k}, \theta_{2\ell}) \in \Theta_S^{(m)}} f^{(m)}(\theta_{1k}, \theta_{2\ell})$ converges to an integral over the line as $m \rightarrow \infty$, it has a limit of zero. Thus, we conclude

$$\lim_{m \rightarrow \infty} d(\widehat{\mu}_\varepsilon^{(m)}, f^{(m)}) = \varepsilon.$$

This completes the proof of Step 1. □

We say that the EAW condition holds for the environment $(\gamma, \Theta^{(m)}, \mu^{(m)})$ if the distribution $\mu^{(m)}$ and the collection of valuations $(\theta_{1k} + \gamma\theta_{2\ell}, \theta_{2\ell} + \gamma\theta_{1k})_{(\theta_{1k}, \theta_{2\ell}) \in \Theta^{(m)}}$ satisfy the EAW condition. Otherwise, we say that the EAW condition fails.

Step 2: For every $\varepsilon > 0$, there exists $\bar{\gamma}_\varepsilon \in (0, 1)$ such that, for all $\gamma \in (\bar{\gamma}_\varepsilon, 1)$ and all $m \in \mathbb{N}$ large enough, the EAW condition fails for the environment $(\gamma, \Theta^{(m)}, \hat{\mu}_\varepsilon^{(m)})$.

Proof. Substituting the linear valuation functions and probability mass function $\hat{\mu}_\varepsilon^{(m)}$ into the EAW condition, we obtain

$$\begin{aligned} & \sum_{k=1}^m \sum_{\ell=k}^m \hat{\mu}_\varepsilon^{(m)}(\theta_{1k}, \theta_{2\ell})(1 - \gamma)(\theta_{2\ell} - \theta_{1k}) \\ & \geq \sum_{k=1}^{m-1} \hat{\mu}_\varepsilon^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{1m} - \theta_{1k}) + \sum_{\ell=2}^m \hat{\mu}_\varepsilon^{(m)}(\theta_{11}, \theta_{2\ell})(\theta_{2\ell} - \theta_{21}). \end{aligned} \quad (11)$$

We use the decomposition $\hat{\mu}_\varepsilon^{(m)} = (1 - \varepsilon)f^{(m)} + \varepsilon\mu_S^{(m)}$. Then, both sides can be decomposed into two components: the continuous part (from $f^{(m)}$) and the singular part (from $\mu_S^{(m)}$).

We first consider the left-hand side (LHS) of inequality (11):

$$\begin{aligned} \text{LHS} &= \sum_{k=1}^m \sum_{\ell=k}^m \left[(1 - \varepsilon)f^{(m)}(\theta_{1k}, \theta_{2\ell}) + \varepsilon\mu_S^{(m)}(\theta_{1k}, \theta_{2\ell}) \right] (1 - \gamma)(\theta_{2\ell} - \theta_{1k}) \\ &= (1 - \varepsilon)(1 - \gamma) \sum_{k=1}^m \sum_{\ell=k}^m f^{(m)}(\theta_{1k}, \theta_{2\ell})(\theta_{2\ell} - \theta_{1k}) \\ &\quad + \varepsilon(1 - \gamma) \left[\sum_{\ell=2}^m \mu_S^{(m)}(\theta_{11}, \theta_{2\ell})(\theta_{2\ell} - \theta_{11}) + \sum_{k=2}^{m-1} \mu_S^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{2m} - \theta_{1k}) \right]. \end{aligned}$$

To simplify notation, we define $C^{(m)} = \sum_{k=1}^m \sum_{\ell=k}^m f^{(m)}(\theta_{1k}, \theta_{2\ell})(\theta_{2\ell} - \theta_{1k})$. We continue the computation:

$$\begin{aligned} \text{LHS} &= (1 - \varepsilon)(1 - \gamma)C^{(m)} \\ &\quad + \varepsilon(1 - \gamma) \left[\frac{1}{2(m-1)} \sum_{\ell=2}^{m-1} \frac{\ell-1}{m-1} + \frac{1}{m-1} + \frac{1}{2(m-1)} \sum_{k=2}^{m-1} \left(1 - \frac{k-1}{m-1} \right) \right] \\ &= (1 - \varepsilon)(1 - \gamma)C^{(m)} \\ &\quad + \varepsilon(1 - \gamma) \left[\frac{1}{2(m-1)^2} \sum_{\ell=2}^{m-1} (\ell-1) + \frac{1}{m-1} + \frac{1}{2(m-1)^2} \sum_{k=2}^{m-1} (m-k) \right]. \end{aligned}$$

After rearrangement, we obtain

$$\begin{aligned}
\text{LHS} &= (1 - \varepsilon)(1 - \gamma)C^{(m)} + \varepsilon(1 - \gamma) \left[\frac{2}{2(m-1)^2} \frac{(m-2)(m-1)}{2} + \frac{1}{m-1} \right] \\
&= (1 - \varepsilon)(1 - \gamma)C^{(m)} + \varepsilon(1 - \gamma) \left[\frac{m-2}{2(m-1)} + \frac{1}{m-1} \right] \\
&= (1 - \varepsilon)(1 - \gamma)C^{(m)} + \varepsilon(1 - \gamma) \frac{m}{2(m-1)}.
\end{aligned}$$

We proceed to compute the right-hand side (RHS) of inequality (11). The RHS can also be decomposed into two components: the continuous part (from $f^{(m)}$) and the singular part (from $\mu_S^{(m)}$). The RHS is thus rewritten as

$$\begin{aligned}
\text{RHS} &= \sum_{k=1}^{m-1} \left[(1 - \varepsilon)f^{(m)}(\theta_{1k}, \theta_{2m}) + \varepsilon\mu_S^{(m)}(\theta_{1k}, \theta_{2m}) \right] (\theta_{1m} - \theta_{1k}) \\
&\quad + \sum_{\ell=2}^m \left[(1 - \varepsilon)f^{(m)}(\theta_{11}, \theta_{2\ell}) + \varepsilon\mu_S^{(m)}(\theta_{11}, \theta_{2\ell}) \right] (\theta_{2\ell} - \theta_{21}) \\
&= (1 - \varepsilon) \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{1m} - \theta_{1k}) \\
&\quad + \varepsilon \left[\sum_{k=1}^{m-1} \mu_S^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{1m} - \theta_{1k}) + \sum_{\ell=2}^m \mu_S^{(m)}(\theta_{11}, \theta_{2\ell})(\theta_{2\ell} - \theta_{21}) \right],
\end{aligned}$$

where the second equality follows because $f^{(m)}(\theta_{11}, \theta_{2\ell}) = 0$ for all $\ell \in \{2, \dots, m\}$. Recall the singular distribution $\mu_S^{(m)}$: $\mu_S^{(m)}(\theta_{11}, \theta_{2m}) = 1/(m-1)$ and $\mu_S^{(m)}(\theta_{1k}, \theta_{2m}) = \mu_S^{(m)}(\theta_{11}, \theta_{2\ell}) = 1/(2(m-1))$ for each $k \in \{2, \dots, m-1\}$ and each $\ell \in \{2, \dots, m-1\}$. Substituting it into the above expression, we obtain

$$\begin{aligned}
\text{RHS} &= (1 - \varepsilon) \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{1m} - \theta_{1k}) \\
&\quad + \varepsilon \left[\frac{1}{m-1} + \sum_{k=2}^{m-1} \frac{1}{2(m-1)} \frac{m-k}{m-1} + \sum_{\ell=2}^{m-1} \frac{1}{2(m-1)} \frac{\ell-1}{m-1} + \frac{1}{m-1} \right] \\
&= (1 - \varepsilon) \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{2m} - \theta_{1k}) \\
&\quad + \varepsilon \left[\frac{2}{m-1} + \frac{1}{2(m-1)^2} \sum_{k=2}^{m-1} (m-k) + \frac{1}{2(m-1)^2} \sum_{\ell=2}^{m-1} (\ell-1) \right] \\
&= (1 - \varepsilon) \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{2m} - \theta_{1k}) + \varepsilon \left[\frac{2}{m-1} + \frac{2}{2(m-1)^2} \frac{(m-2)(m-1)}{2} \right].
\end{aligned}$$

To simplify notation, we define $\widehat{C}^{(m)} = \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{2m} - \theta_{1k})$. We continue the computation:

$$\text{RHS} = (1 - \varepsilon)\widehat{C}^{(m)} + \varepsilon \left[\frac{2}{m-1} + \frac{m-2}{2(m-1)} \right] = (1 - \varepsilon)\widehat{C}^{(m)} + \frac{(m+2)\varepsilon}{2(m-1)}.$$

To summarize, the EAW condition requires

$$\begin{aligned} (1 - \varepsilon)(1 - \gamma)C^{(m)} + \varepsilon(1 - \gamma)\frac{m}{2(m-1)} &\geq (1 - \varepsilon)\widehat{C}^{(m)} + \frac{(m+2)\varepsilon}{2(m-1)} \\ \Rightarrow 1 - \gamma &\geq \frac{(1 - \varepsilon)\widehat{C}^{(m)} + \frac{(m+2)\varepsilon}{2(m-1)}}{(1 - \varepsilon)C^{(m)} + \frac{m\varepsilon}{2(m-1)}} \\ \Rightarrow \gamma &\leq \frac{(1 - \varepsilon) \left(C^{(m)} - \widehat{C}^{(m)} \right) - \frac{2\varepsilon}{2(m-1)}}{(1 - \varepsilon)C^{(m)} + \frac{m\varepsilon}{2(m-1)}}. \end{aligned}$$

To simplify notation, we define

$$\bar{\gamma}(\varepsilon, m) = \frac{(1 - \varepsilon) \left(C^{(m)} - \widehat{C}^{(m)} \right) - \frac{2\varepsilon}{2(m-1)}}{(1 - \varepsilon)C^{(m)} + \frac{m\varepsilon}{2(m-1)}}.$$

Then, the EAW condition is reduced to $\gamma \leq \bar{\gamma}(\varepsilon, m)$. Taking the limit as $m \rightarrow \infty$, we have

$$\bar{\gamma}_\varepsilon = \lim_{m \rightarrow \infty} \bar{\gamma}(\varepsilon, m) = \lim_{m \rightarrow \infty} \frac{2(1 - \varepsilon)(C^{(m)} - \widehat{C}^{(m)}) - \frac{2\varepsilon}{m-1}}{2(1 - \varepsilon)C^{(m)} + \frac{m\varepsilon}{m-1}} = \frac{2(1 - \varepsilon) \lim_{m \rightarrow \infty} (C^{(m)} - \widehat{C}^{(m)})}{2(1 - \varepsilon) \lim_{m \rightarrow \infty} C^{(m)} + \varepsilon}.$$

Observe that

$$\lim_{m \rightarrow \infty} \widehat{C}^{(m)} = \lim_{m \rightarrow \infty} \sum_{k=2}^{m-1} f^{(m)}(\theta_{1k}, \theta_{2m})(\theta_{2m} - \theta_{1k}) = \int_0^1 (1 - \theta_1) f(\theta_1, 1) d\theta_1 \geq 0,$$

where the weak inequality holds because $1 - \theta_1 \geq 0$ for all $\theta_1 \in [0, 1]$. On the other hand, we also have

$$\lim_{m \rightarrow \infty} \widehat{C}^{(m)} = \int_0^1 (1 - \theta_1) f(\theta_1, 1) d\theta_1 \leq \int_0^1 f(\theta_1, 1) d\theta_1 = 0,$$

where the weak inequality follows because $1 - \theta_1 \leq 1$ for all $\theta_1 \in [0, 1]$. Therefore, we obtain

$$\lim_{m \rightarrow \infty} \widehat{C}^{(m)} = 0.$$

Furthermore, since

$$\lim_{m \rightarrow \infty} C^{(m)} = \lim_{m \rightarrow \infty} \sum_{k=1}^m \sum_{\ell=k}^m f^{(m)}(\theta_{1k}, \theta_{2\ell})(\theta_{2\ell} - \theta_{1k}) > 0,$$

we have $\bar{\gamma}_\varepsilon \in (0, 1)$. We conclude that for all $\gamma \in (\bar{\gamma}_\varepsilon, 1)$ and all $m \in \mathbb{N}$ large enough, the EAW condition fails for $(\gamma, \Theta^{(m)}, \hat{\mu}_\varepsilon^{(m)})$. This completes the proof of Step 2. $\square$

Since the EAW condition is a necessary condition for the existence of desired mechanisms satisfying NTNP in any finite environment, the impossibility of locally robust implementation of efficient bilateral trade is established. $\square$

References

- [1] Bergemann, D. and S. Morris, “Robust Implementation in Direct Mechanisms,” *The Review of Economic Studies*, vol. 76, (2009) 1175-1204.
- [2] Bewley, T.F., *Knightian Decision Theory: Part I*, Discussion paper, Cowles Foundation, 1986.
- [3] B rgers, T., *An Introduction to the Theory of Mechanism Design*, Oxford University Press, 2015.
- [4] B rgers, T., and P. Norman, “A Note on Budget Balance under Interim Participation Constraints: the Case of Independent Types,” *Economic Theory*, vol. 39, (2009), 477-489.
- [5] Cr mer, J. and R.P. McLean, “Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent,” *Econometrica*, vol. 53, (1985), 345-361.
- [6] Cr mer, J. and R.P. McLean, “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions,” *Econometrica*, vol. 56, (1988), 1247-1257.
- [7] Fieseler, K., T. Kittsteiner, and B. Moldovanu, “Partnerships, Lemons, and Efficient Trade,” *Journal of Economic Theory*, vol. 113, (2003), 223-234.
- [8] Garratt, R. and M. Pycia, “Efficient Bilateral Trade,” *Working Paper*, September 2025.
- [9] Gresik, T.A., “Efficient Bilateral Trade with Statistically Dependent Beliefs,” *Journal of Economic Theory*, vol. 53, (1991a), 199-205.

- [10] Gresik, T.A., “Ex Ante Incentive Efficient Trading Mechanisms without the Private Valuation Restriction,” *Journal of Economic Theory*, vol. 55, (1991b), 41-63.
- [11] Kosenok, G. and S. Severinov, “Individually Rational, Budget-balanced Mechanisms and Allocation of Surplus,” *Journal of Economic Theory*, vol. 140, (2008), 126-161.
- [12] Krishna, V. and M. Perry, “Efficient Mechanism Design,” Working Paper, (2000).
- [13] Kunimoto, T., and C. Zhang, “Efficient Bilateral Trade via Two-Stage Mechanisms: Comparison between One-Sided and Two-Sided Asymmetric Information Environments,” *Journal of Mathematical Economics*, vol. 101, (2022), 102714.
- [14] Kunimoto, T., and C. Zhang, “The Interplay of Interdependence and Correlation in Bilateral Trade,” *Working Paper*, 2024.
- [15] Lopomo, G., L. Rigotti, and C. Shannon, “Uncertainty and Robustness of Surplus Extraction,” *Journal of Economic Theory*, vol. 199, (2022), 105088.
- [16] Makowski, L. and C. Mezzetti, “Bayesian and Weakly Robust First Best Mechanisms: Characterizations,” *Journal of Economic Theory*, vol. 64, (1994), 500-519.
- [17] Matsuo, T., “On Incentive Compatible, Individually Rational, and Ex Post Efficient Mechanisms for Bilateral Trading,” *Journal of Economic theory*, vol. 49, (1989), 189-194.
- [18] Myerson, R.B. and M.A. Satterthwaite, “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, vol. 29, (1983), 265-281.
- [19] Williams, S., “A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms,” *Economic Theory*, vol. 14, (1999), 155-180.