Cooperative Analysis of Incomplete Information

Qingmin Liu*

May 30, 2023

Abstract

We propose a theory of cooperative games with incomplete information. The theory concerns stable interactions that cannot be undermined by coalitions and is built on a criterion of rational counterfactual reasoning, which requires that in every counterfactual scenario of coalitional deviation, every individual player formulate a belief and act optimally, and in doing so they collectively prevent the counterfactual scenario from actualizing. We equip the criterion with *weak consistency* and *strong consistency* to reflect the alignment of players' beliefs and incentives in deviating coalitions, and demonstrate their implications through two applications. We identify a condition of *comonotonic differences* that preserves the efficiency of stable matching with incomplete information, where Tarski's fixed point theorem is a useful tool. We show that mutual costly signaling unravels outcome-relevant incomplete information in networks.

1 Introduction

Cooperative solution concepts, such as the core and pairwise stability, find a wide range of applications, for example in labor markets, school choices, social networks, family

*Department of Economics, Columbia University. E-mail: ql2177@columbia.edu. I thank (in chronological order) Tianhao Liu, Balazs Szentes, Benny Moldovanu, Rajiv Vohra, Roberto Serrano, Mike Borns, Jay Lu, Kevin Zollman, Françoise Forges, Sushil Bikhchandani, Eric Maskin, Bentley MacLeod, Yeon-Koo Che, Navin Kartik, Steven Stern, Eran Shmaya, and Pradeep Dubey for helpful comments and discussions. The paper was also improved by feedback from seminars and conferences at Brown, Collegio Carlo Alberto, NUS, UCL, UCLA, Columbia, and Stony Brook.

economics, international relations, and voting. They serve as descriptive theories or normative criteria for analyzing markets and games. Quite remarkably, even though both the *cooperative* framework of *complete* information and the *non-cooperative* framework of *incomplete* information are richly developed, cooperative theories of incomplete information have not received adequate attention; "this area is to this day fraught with unresolved conceptual difficulties" (Aumann and Heifetz, 2002); cooperative game-theoretic analysis of practical economic applications with incomplete information is a virtually uncharted territory. This under-exploration is especially noteworthy, given the methodological comparative advantages of cooperative models for complex strategic interactions on the one hand, and the practical prevalence of incomplete information on the other.[1]

	Complete Information	Incomplete Information
Non-Cooperative	✓	✓
Cooperative	✓	?

This paper presents a general approach to cooperative analysis of incomplete information and explores its applications. Focusing on the core and stability, our departure from the existing literature is both conceptual and applicational.

The foremost conceptual question of cooperative analysis of incomplete information is to model how players make inferences about one another and form beliefs without relying on ad hoc assumptions about individual strategies and strategic interactions. The literature has experienced significant theoretical advancements since the breakthrough of Wilson (1978).[2] At the risk of oversimplification, one can describe the dominant paradigm as modeling the processes through which players form coalitions and exchange information, often reintroducing mechanism design or other non-cooperative elements back to a cooperative framework. This approach naturally arises from recognizing each scenario of coalitional deviation as a complex multiple-player game.

Our methodology is different. The advantages of cooperative analysis come from

[1] Non-cooperative analysis is usually sensitive to assumptions about strategic interactions (e.g., strategy spaces, dynamics, order of play, information, etc. about which the analysts often have limited knowledge) and can be intractable when the underlying interactions are complex or involve many parties, see, e.g., Fisher (1991) for an intriguing discussion in the context of industrial organization.

[2] See, e.g., Holmström and Myerson (1983) and Forges, Minelli, and Vohra (2002).

sidestepping the complexity and flexibility of non-cooperative modeling. To this end, we approach the problem in reverse. The starting point is the realization that rational Bayesian players must arrive at some probabilistic beliefs in every possible scenario, regardless of how they interact and make inferences. Therefore, a cooperative solution to a strategic problem should explicitly articulate players' beliefs in each counterfactual scenario of coalitional deviation and ensure that each player acts optimally against the system of beliefs, which prevents the coalitional deviation from taking place (so the counterfactual scenario remains counterfactual). This is what we refer to as rational counterfactual reasoning.

After setting the stage, the most important conceptual question is to identify meaningful constraints on beliefs and outcomes that can have sufficient prediction power. We propose two types of constraints, *weak consistency* and *strong consistency*, to reflect the alignment of deviating players' beliefs and incentives in a deviating coalition. The difference is subtle but crucial. A scenario of deviation is relevant for a player only when other players (if any) in the same coalition expect to benefit from the deviation, as per the system of beliefs, and therefore willingly participate in it. Weak consistency requires that players' beliefs in each scenario of deviation be Bayesian consistent with their peers' willingness to deviate, whenever possible. Strong consistency and its variants require that players' beliefs "mutually reassure" others' deviation—meaning, their willingness to deviate is dependent on one another. This mutual reassurance can generate more information that will refine the original belief system. Each counterfactual scenario of deviation is indeed a multiple-player game, but the formulation does not rely on non-cooperative processes of information exchange or belief formation and hence adheres to the principles of cooperative game theory, taking a reduced-form approach to complex games.

The general model captures applications such as labor markets, marriage markets, trading networks, exchanges, etc. In these contexts, there isn't a planning stage, and re-contracting opportunities arise after outcomes are observed: divorces happen in existing marriages, new connections are established and old connections are removed within existing networks, firms adjust their existing workforce with hiring and layoff decisions, etc.[3] This feature is not captured by the traditional trichotomy of ex ante, interim,

[3]Even in markets that involve some degree of planning, participants may not adhere to their initial

and ex post frameworks, or by the analysis of interim deviations from ex ante agreements (i.e., durable decision rules) in Holmström and Myerson (1983), highlighting the need of a new analytical framework.[4] It is worth noting that the prevailing outcomes (marriages, connections, contracts, etc.) inform about the incentives and disincentives for deviating from them, and this information in turn shapes the outcomes in a feedback loop. This information perspective is another difference between our approach and familiar cooperative or non-cooperative equilibrium models, which opens up a wealth of new applications and sheds new light on well-studied ones. Markets with adverse selection are an example in point. For instance, if a separating contract prevails in an insurance market, policyholders will reveal their types. Competing insurance companies or new entrants will then take advantage of this information and offer full coverage contracts tailored to specific types, thus destabilizing the separating outcome. Therefore, the framework of Rothschild and Stiglitz (1976) is not suitable for analyzing stable insurance markets. These familiar applications, previously analyzed in non-cooperative models, warrant further investigation. We foresee that the ideas developed in our cooperative framework of incomplete information will become valuable in a wide range of applications.

We study two concrete applications in depth, matching and networks with incomplete information, which are of significant research interest but have been limited in progress due to their complexity. While these applications often involve many players, a typical strategic interaction is unilateral or bilateral. The recent proliferation of research in this area, which mostly assumes complete information, attests to their versatility and prevalence.

We assume that all players possess private information in the two-sided matching application. In a Bayesian framework, two-sided incomplete-information matching games raise new conceptual questions that are not present in the special case of one-sided incomplete information settings, such as the one studied in Liu (2020), because they necessitate considering players' mutual inferences from each other's incentives and

agreements and may enter into side agreements after seeing the realized outcomes.

[4]Green and Laffont (1987) refine the notion of implementability and Forges (1994) studies modified notions of Pareto efficiency with this "posterior" re-contracting possibility. In contrast, our focus is on a solution concept that captures stabilized market interactions and a self-reinforcing information loop plays a crucial role in shaping stable outcomes.

information that jointly determine the formation of beliefs. Restricting to one-sided incomplete information significantly limits potential applications.[5] This difference is further manifested in the distinction between the concepts of weak consistency and strong consistency (Theorem 5). Readers who are familiar with non-cooperative models of bargaining and signaling will find the distinction we are making here quite obvious. These existing models are typically restricted to one-sided incomplete information in two-player games; however, multiple-player games with multiple-sided incomplete information are substantively different and often intractable, significantly limiting the study of many economic applications. This paper offers a tractable cooperative approach.

We identify conditions, for weak and strong consistencies, respectively, under which all stable matchings with transfers maximize the total surplus from an outside observer's perspective. Specifically, for weak consistency, we show that the interdependence of players' payoffs becomes relevant. For strong consistency, we demonstrate that a new condition, called "comonotonic differences," is crucial for restoring efficiency of stability, a duality that always hold under complete information. This concept of comonotonic differences says that, for any pair of deviating players, their types can be reordered in a way that aligns their incentives to deviate. Mathematically, the condition ensures the existence of a fixed-point pair of sets of types ("mutually reassuring set") that defines strong consistency. Interestingly, Tarski's fixed point theorem once again becomes a useful tool in matching, but for a very different reason from complete-information games. We consider efficient matching from an outside observer's perspective. The criterion is useful because the analysts are not directly involved in the economic activity and have no direct access to insiders' private information. They need a theoretical framework to understand and interpret matching data, perform evaluations, and recommend policy interventions. The concepts of stability, weak consistency, and strong consistency, and the criterion of efficiency serve as such a framework.

[5]There is a small growing literature on matching with incomplete information. Liu, Mailath, Postlewaite, and Samuelson (2014) devise an iterative concept of stability for matching with one-sided incomplete information, leaving open the determination of beliefs that should be endogenous. Bikhchandani (2017) makes exogenous assumptions on beliefs in the iterative process. Chen and Hu (2023) extend the iterative approach to two-sided incomplete information, where the resulting solution is permissive even under strong assumptions. Liu (2020) proposes a "Kreps–Wilson program" where solution concepts must specify correct on-path beliefs and consistent off-path beliefs. The present paper can also be seen as an advancement of this program.

The second application is networks. Despite the many applications of creating and maintaining relationships between individuals or organizations with incomplete information, network modeling remains primarily focused on complete information. We make two additional assumptions to capture plausible features of stabilized network structures and link formations. First, directly connected players in an existing network observe each other's characteristics, although players who are not connected or indirectly connected may not. Second, before establishing a new connection, players can engage in mutual signaling activities, such as transfers to third parties in or out of the network, investment in education and physical characteristics, conspicuous consumption, virtue signaling, etc. We identify conditions under which strongly consistent and stable networks under incomplete information must be state-by-state complete-information stable. However, it is important to note that although network structures unravels, uncertainty may persist within incomplete-information stable networks. The result helps in understanding how incomplete information affects network formation, how the existing models that assume complete information are robust in terms of final distributions of connections, as well as the scope of policies and interventions that could influence the stabilized network structures under uncertainty.

This paper leaves several important issues open for further investigation. The first is non-cooperative implementation of the cooperative concepts we developed, specifically an incomplete-information version of the Nash program. The second issue is concerned with alternative cooperative concepts. The third is to study economic applications that involve larger deviation coalitions. The paper is unfortunately too short to cover all the interesting topics it opens up, but might well be too long for a reader to engage with comfortably. We provide some additional discussion in Section 8.

The rest of the paper is organized as follows. Section 2 defines coalitional games with incomplete information and Section 3 describes several familiar applications. Section 4 formulates belief systems and the notion of stability. Section 5 formulates weak and strong consistencies, which are the main analytical tools. Sections 6 and 7 study applications of matching and networks, respectively. The appendix contains omitted proofs and additional materials.

2 Games with Incomplete Information

Let N be a finite set of players. Player $n \in N$ privately knows his own type $\theta_n \in \Theta_n$, where Θ_n is finite. The state space is $\Theta = \prod_{n \in N} \Theta_n$. Let $\beta_n^0 \in \Delta(\Theta)$ be player n's prior belief, which is assumed to have a full support for notational simplicity.

Let Z be the set of feasible **outcomes**. We do not restrict Z to be finite. Each player n has a von Neumann–Morgenstern utility function $u_n : Z \times \Theta \to \mathbb{R}$. Let $\mathbb{S}$ be a collection of non-empty subsets of N. Each set $S \in \mathbb{S}$ is a **coalition**. We write $\Theta_S = \prod_{n \in S} \Theta_n$ and $\Theta_{-S} = \prod_{n \notin S} \Theta_n$. For each coalition $S \in \mathbb{S}$, let $d(S, z) \subset Z$ be the set of **feasible outcomes** which the coalition S can collectively effectuate from z.We do not formulate a cooperative game in characteristic-function form, which is not convenient for an incomplete-information setting as it will become clear.

A game with incomplete information is $\Gamma = (N, Z, \mathbb{S}, d, (u_n, \Theta_n, \beta_n^0)_{n \in N})$. Denote by $\Gamma^\theta = (N, Z, \mathbb{S}, d, (u_n(\cdot, \theta))_{n \in N})$ the special case of complete information when $\theta \in \Theta$ is commonly known. The natural solution concept for Γ^θ is as follows.

Definition 1. *An outcome $z \in Z$ is a **stable outcome** of Γ^θ if there do not exist $S \in \mathbb{S}$ and $z' \in d(S, z)$ such that $u_n(z', \theta) > u_n(z, \theta)$ for all $n \in S$.*

This widely-applied concept requires a leap of faith, as it is agnostic on how a coalition S is formed, how the coalition agrees on the alternative z', and how a stable outcome z is achieved. However, the simplicity is precisely its advantage. We shall extend this concept to incomplete information, maintaining the same kind of simplification.

3 Examples

The general model encompasses a wide range of problems as special cases, and the complete-information versions of these problems are classic and well understood.

Two-Sided Matching. The set of players is $N = I \cup J$, where $I \cap J = \emptyset$. A *matching outcome* is $z = (\mu, \tau)$, where $\mu = (\mu_n)_{n \in N}$ assigns a partner to each player $n \in N$ s.t.

$$\begin{aligned} &\text{(i)} \ \ \mu_n = m \text{ iff } \mu_m = n, \\ &\text{(ii)} \ \ \mu_n \in J \cup \{n\} \text{ if } n \in I, \\ &\text{(iii)} \ \ \mu_n \in I \cup \{n\} \text{ if } n \in J, \end{aligned}$$

and $\tau = (\tau_n)_{n\in N}$ specifies a transfer $\tau_n \in \mathbb{R}$ to each player n such that $\tau_n + \tau_{\mu_n} = 0$. If $\mu_n = n$, player n is *unmatched*. We write $z_n = (\mu_n, \tau_n)$ as player n's matching outcome. Let Z be the set of all matching outcomes. A coalition consists of a single player or a pair:

$$\mathbb{S} = \{\{i\}, \{j\}, \{i,j\} : i \in I, j \in J\}. \tag{3.1}$$

For any $z = (\mu, \tau)$, any individual player n can abandon his partner (if he is matched under z) to stay alone, i.e.,

$$d(\{n\}, z) = \left\{ z' \in Z : \begin{array}{ll} z'_m = (m, 0), & \text{if } m \in \{n, \mu_n\} \\ z'_m = z_m, & \text{otherwise} \end{array} \right\},$$

and any pair $\{i,j\} \in \mathbb{S}$ can match with each other for any transfers, leaving their original partners unmatched, i.e.,

$$d(\{i,j\}, z) = \left\{ z' = (\mu', \tau') \in Z : \begin{array}{ll} \mu'_i = j & \\ \mu'_n = n, & \text{if } n \notin \{i,j\} \text{ and } \mu_n \in \{i,j\} \\ z'_n = z_n, & \text{otherwise} \end{array} \right\}.$$

For applications, it is often convenient to assume that a player's payoff depends only on his own matches and is linear in transfers, i.e., $u_n(z, \theta) = v_n(\mu_n, \theta) + \tau_n$ for some function v_n. In the special case of complete information, Definition 1 is the familiar concept of pairwise stability of Shapley and Shubik (1971) and Crawford and Knoer (1981). If we restrict $\tau_n \equiv 0$, the model reduces to a marriage problem and the special case of complete information is studied by Gale and Shapley (1962). □

Network Formation. A *network* on N is $z = (z_n)_{n\in N}$, where $z_n \subset N$ is player n's neighbors that satisfies

$$\begin{array}{l} \text{(i) } n \in z_n \text{ for all } n \in N; \\ \text{(ii) } m \in z_n \text{ iff } n \in z_m, \text{ for all } m, n \in N. \end{array} \tag{3.2}$$

Condition (i) is for notational convenience only. Let Z be the set of all networks on N, and let $\mathbb{S}$ be the collection of singleton and doubleton subsets of N. Any player i can unilaterally remove any of his neighbors in a network $z \in Z$, i.e.,

$$d(\{i\}, z) = \left\{ (z'_n)_{n\in N} \in Z : \begin{array}{lll} z'_n \subset z_n, & & \text{if } n = i \\ z_n \backslash \{i\} \subset z'_n \subset z_n, & & \text{if } n \in z_i \backslash \{i\} \\ z'_n = z_n, & & \text{otherwise} \end{array} \right\}.$$

It takes two distinct players $\{i,j\} \in \mathbb{S}$ to establish a new link in $z \in Z$, i.e.,

$$d(\{i,j\},z) = \left\{ (z'_n)_{n\in N} \in Z : \begin{array}{ll} z'_n = z_n \cup \{i,j\}, & \text{if } n \in \{i,j\} \\ z'_n = z_n, & \text{otherwise} \end{array} \right\}.$$

In the special case of complete information, Definition 1 is the concept of pairwise stability of Jackson and Wolinsky (1996). □

Exchange Economy. Each player $n \in N$ has an endowment $e_n \in \mathbb{R}_+^{|K|}$, where K is a finite set of commodities. Let $z_n \in \mathbb{R}_+^{|K|}$ be player n's consumption bundle, and let $z^k \in \mathbb{R}_+^{|N|}$ be all players' consumption of commodity k. An outcome of the economy is $z = (z_n)_{n\in N} = (z^k)_{k\in K}$. The set of feasible coalitions $\mathbb{S}$ consists of all non-empty subsets of N, and for any $S \in \mathbb{S}$, we write $1_S = (x_n)_{n\in N}$ as a vector of 0's and 1's such that $x_n = 1$ iff $n \in S$. The set of feasible outcomes of this economy is

$$Z = \left\{ z \in \mathbb{R}_+^{|K||N|} : 1_N \cdot z^k \leq 1_N \cdot e^k \text{ for all } k \in K \right\}.$$

The set of feasible outcomes for a coalition $S \in \mathbb{S}$ is

$$d(S,z) = \left\{ z' \in Z : 1_S \cdot z^k \leq 1_S \cdot e^k \text{ for all } k \in K \right\}.$$

Thus, players in S can trade exclusively among themselves and leave players in $N \backslash S$ to trade among themselves. Under complete information, Definition 1 corresponds to the core of the economy. The incomplete information is about preference uncertainty (as contrast to endowment uncertainty). □

Competitive Price Equilibrium. In addition to e_n, z_n and z^k as in the exchange economy model, let $z_0 \in \mathbb{R}_+^{|K|}$ be a price vector. A market outcome is $z = (z_n)_{n\in N\cup\{0\}} \in \mathbb{R}_+^{|K|(|N|+1)}$. The set of feasible outcomes is

$$Z = \left\{ z = (z_n)_{n\in N\cup\{0\}} \in \mathbb{R}_+^{|K|(|N|+1)} : \begin{array}{ll} 1_N \cdot z^k \leq 1_N \cdot e^k & \text{for all } k \in K \\ z_0 \cdot z_n \leq z_0 \cdot e_n & \text{for all } n \in N \end{array} \right\}.$$

Player n's utility depends on z only through z_n. Only unilateral deviations are allowed, so $\mathbb{S} := \{\{n\} : n \in N\}$ and $d(\{n\},z) = \{z' \in Z : z'_0 = z_0\}$. The interpretation is that player n, a price taker, can unilaterally change his consumption from z_n to z'_n as long as the outcome z' remains feasible under price z_0. The restriction that $d(\{n\},z) \subset Z$ says that consumption must be physically possible (i.e., consumption is constrained by the endowments of the economy) and affordable (it is still mediated by the price system). In the traditional formulation of a Walrasian equilibrium, the only constraint

for player n's deviation is the budget constraint $z_0 \cdot z'_n \leq z_0 \cdot e_n$; the endowment constraint is considered for the equilibrium consumption, but *not* for the deviation, so n's off-equilibrium consumption can exceed what the economy offers. Definition 1 defines a notion of price equilibrium that is weaker than the Walrasian equilibrium. □

4 Rational Counterfactual Reasoning and Stability

Rational players should evaluate all possible counterfactual scenarios of deviation. Being Bayesian, they should also ascribe probabilistic beliefs to uncertainties at all counterfactual scenarios, use Bayes' rule whenever possible, and act optimally given their beliefs in each counterfactual scenario. In any scenario of coalitional deviation of a putative stable situation, some player in the coalition should find it optimal *not* to engage in the deviation, thus enforcing the counterfactuality of the scenario. We shall formalize these ideas.

The relationship between the underlying uncertainties and "stabilized" outcomes is described by a function $\pi : \Theta \to Z$. We refer to π as a **playout** of the game.[6] If the game is *played out* according to π, then $\pi(\Theta) = \{\pi(\theta) : \theta \in \Theta\} \subset Z$ is the set of outcomes that actualize. Let

$$\Delta_\pi = \{(S, z, z') : \ S \in \mathbb{S},\ z \in \pi(\Theta),\ z' \in d(S, z)\}$$

be the set of **deviations** associated with the playout π. Let

$$\Sigma_\pi = \{(S, z, z', \theta) : (S, z, z') \in \Delta_\pi,\ z = \pi(\theta)\}$$

be the **scenarios of deviations** of π. A scenario of deviation $\sigma = (S, z, z', \theta) \in \Sigma_\pi$ indicates that a coalition S can deviate to z' from z at state θ. But player n's **perception** of σ is $\sigma_n = (S, z, z', \theta_n)$ since he does not know θ_{-n}. Let $\Sigma_{\pi,n}$ be player n's **perceived scenarios of deviations**, i.e.,

$$\Sigma_{\pi,n} = \{(S, z, z', \theta_n) : n \in S,\ (S, z, z', \theta_n, \theta_{-n}) \in \Sigma_\pi \text{ for some } \theta_{-n} \in \Theta_{-n}\}.$$

Let $\pi_n^{-1}(z)$ be the projection of $\pi^{-1}(z)$ on Θ_n. Then,

$$\Sigma_{\pi,n} = \{(S, z, z', \theta_n) : n \in S, (S, z, z') \in \Delta_\pi, \theta_n \in \pi_n^{-1}(z)\}.$$

A Bayesian player has a belief in each scenario. Let $\beta_n : \Sigma_{\pi,n} \to \Delta(\Theta)$ be a mapping

[6] A version of "correlated stability" through the playout $\pi : \Theta \times T \to Z$ can be defined analogously on an enlarged state space $\Theta \times T$.

that specifies a belief $\beta_n(\sigma_n)$ for player n at each of his perceived scenarios of deviation $\sigma_n \in \Sigma_{\pi,n}$.

Definition 2. *A **belief system** for a playout π is $\beta = (\beta_n)_{n\in N}$ such that for any $\sigma_n = (S, z, z', \theta_n) \in \Sigma_{\pi,n}$, we have*

(i) *self-recognition: $\beta_n(\sigma_n)(\{\theta_n\} \times \Theta_{-n}) = 1$, and*

(ii) *knowledge of playout: $\beta_n(\sigma_n)(\pi^{-1}(z)) = 1$.*

*We refer to a playout–belief pair (π, β) of a game Γ as an **assessment** of the game.*

The terminology of assessment is borrowed from Kreps and Wilson (1982), but it should be noted that the formulation of a cooperative game does not involve a game tree or individual strategies, and hence making belief restrictions will be our main conceptual and methodological tasks.

Each scenario of deviation corresponds to a complex multiple-player interaction, but since beliefs are specified explicitly, we no longer need to formulate the strategic interaction in a non-cooperative way. This is the advantage of the approach.

Definition 3. *A scenario of deviation $\sigma = (S, z, z', \theta) \in \Sigma_\pi$ is a **viable scenario** with respect to the belief system β if*

$$\mathbf{E}_{\beta_n(\sigma_n)}(u_n(z', \cdot)) > \mathbf{E}_{\beta_n(\sigma_n)}(u_n(z, \cdot))$$

*for all $n \in S$, where $\mathbf{E}_{\beta_n(\sigma_n)}$ is the expectation operator with respect to player n's belief $\beta_n(\sigma_n)$ for the perceived scenario σ_n. A deviation $(S, z, z') \in \Delta_\pi$ is a **viable deviation** if there exists some $\theta \in \Theta$ such that $(S, z, z', \theta) \in \Sigma_\pi$ is a viable scenario.*

For a playout π endowed with a belief system β to be "stable," every scenario of coalitional deviation should remain counterfactual.

Definition 4. *An assessment (π, β) is a **stable assessment** if there is no viable scenario of deviation with respect to β. A playout π is a **stable playout** if (π, β) is a stable assessment for some belief system β.*

Under complete information, this new concept is equivalent to the classic one in Definition 1. It requires the same kind of leap of faith: it is agnostic about how a scenario of deviation arises, only requiring that a stable situation withstands all possible

deviations. Additionally, players must form beliefs for each counterfactual scenario, a leap of faith required for the cooperative formulation of incomplete information.

The stability concept is permissive if there are no further restrictions, because the belief system is quite arbitrary except that it must be compatible with deviation incentives as outlined in Definitions 2–4. Nevertheless, it lays the groundwork for cooperative analysis. As a proof of concept, we first demonstrate that the concept does have some restrictions in special cases.

Definition 5. *A game* $\Gamma = (N, Z, \mathbb{S}, d, (u_n, \Theta_n, \beta_n^0)_{n \in N})$ *has* ***essentially private values*** *if for any* $n \in N$ *there exist* $v_n : Z \times \Theta_n \to \mathbb{R}$, $a_n : \Theta \to \mathbb{R}_{++}$, *and* $b_n : \Theta \to \mathbb{R}$ *such that* $u_n(z, \theta) = a_n(\theta) v_n(z, \theta_n) + b_n(\theta)$ *for all* $z \in Z$ *and* $\theta \in \Theta$.

Player n's payoff in a game with essentially private values depends only on θ_n up to a positive affine transformation that can depend on the type profile θ. The following result is obtained by comparing definitions. Its proof is omitted.

Theorem 1. *Suppose that* Γ *has essentially private values. Then* (π, β) *is a stable assessment of* Γ *if and only if* $\pi(\theta)$ *is a stable outcome of* Γ^θ *for all* $\theta \in \Theta$.

5 Weak and Strong Consistency

In this section, we define the key concepts that give the theory teeth. We do this by identifying constraints on beliefs that capture the inferences that players could make. There are many plausible constraints, but we would like to strike a balance between simplicity, predictive power, and interpretability.

5.1 Weak Consistency

What can player n infer from his perceived scenario of deviation (S, z, z', θ_n)? We should realize that a player's belief becomes relevant only when it leads to consequential decisions. A player's decision at a scenario of deviation is relevant if and only if his opponents join the deviation—if some of his opponents does not join this deviation, the player's decision and belief are irrelevant (this is perhaps the least amount of inference the player can and should make, without making a further leap of faith, from both the player's and the analyst's perspectives). So the player's belief that is relevant for

the analysis should be conditioned on the set of states in which his opponents join the deviation. Furthermore, if the player is the only one involved in the deviation, $S = \{n\}$, his belief should not vary based on their contemplation of different alternatives $z' \neq z''$.

We formulate the notion of "weak consistency" that captures the ideas above. For any deviation $(S, z, z') \in \Delta_\pi$, let $D_n^\beta(S, z, z')$ be player n's **deviating set**, the set of types with which player $n \in S$ benefits from the deviation according to the belief system β, i.e.,

$$D_n^\beta(S, z, z') := \Big\{\theta_n \in \pi_n^{-1}(z) : \mathbf{E}_{\beta_n(S,z,z',\theta_n)}(u_n(z', \cdot)) > \mathbf{E}_{\beta_n(S,z,z',\theta_n)}(u_n(z, \cdot))\Big\}. \tag{5.1}$$

This set can be empty. We write

$$D_{S\setminus\{n\}}^\beta(S, z, z') := \prod\nolimits_{m\in S\setminus\{n\}} D_m^\beta(S, z, z')$$

as the Cartesian product of the deviating sets of player n's opponents. Player n's decision is relevant in his perceived scenario of deviation (S, z, z', θ_n) if and only if his opponents participate in the deviation. Therefore, we require that $\beta_n(S, z, z', \theta_n)$ be updated from the prior belief conditional on $D_{S\setminus\{n\}}^\beta(S, z, z')$ and player n's type, θ_n:

$$\beta_n(S, z, z', \theta_n)(\cdot) = \beta_n^0\Big(\cdot|(\{\theta_n\} \times D_{S\setminus\{n\}}^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z)\Big). \tag{5.2}$$

So, player n assigns positive probability to a state θ only when his opponents all benefit from the deviation in state θ. Bayes' rule has no restriction if not all of player n's opponents participate in the deviation, i.e., $D_{S\setminus\{n\}}^\beta(S, z, z') = \emptyset$, but Definition 2 imposes a support constraint on the conditional probability measure through (5.2). Notice that (5.2) is *not* a definition of beliefs, since β is given and appears on both sides of the equation. It is a *fixed-point* property for the given assessment (π, β) to satisfy.

Definition 6. *An assessment (π, β) is **weakly consistent** if (5.2) holds for all players in all their perceived scenarios of deviations.*

Remark 1. If $S = \{n\}$, then the nullary Cartesian product $D_{S\setminus\{n\}}^\beta(S, z, z') = \{\emptyset\}$. Weak consistency implies that

$$\beta_n(S, z, z', \theta_n)(\cdot) = \beta_n^0\Big(\cdot|(\{\theta_n\} \times \Theta_{-n}) \cap \pi^{-1}(z)\Big),$$

which is independent of z'. That is to say, a player does not learn any information from contemplating a unilateral deviation, as one should expect. In this case, player n updates his belief based on his own type θ_n and the observation of outcome z. □

Remark 2. By Definition 3 and the definition of deviating sets, a deviation (S, z, z') is viable if and only if

$$(D_S^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset,$$

where $D_S^\beta(S, z, z')$ is the Cartesian product $\prod_{n \in S} D_n^\beta(S, z, z')$. Weak consistency imposes restrictions even if no deviation is viable because it restricts the beliefs under which viability is evaluated. If $\theta \in (D_S^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset$, then everyone in S gains from the deviation (S, z, z'). Weak consistency says that, in state θ, everyone's belief is supported by $(D_S^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z)$, so everyone in S believes that everyone in S gains from the deviation. Not limited to first-order beliefs as it may appear, weak consistency implies that, in θ, everyone in S believes that everyone in S believes that everyone in S gains from the deviation, ad infinitum. In epistemic jargon, $(D_S^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z)$ is a "self-evident" common knowledge event.[7] Viability alone, as in Definition 3, does not satisfy this common knowledge property. □

The following example demonstrates the restriction of weak consistency.

Example 1. Consider a partnership game with two players $N = \{1, 2\}$. Each player has two types: $\Theta_n = \{\theta_n^1, \theta_n^2\}$. There is a uniform common prior $\beta^0 \in \Delta(\Theta)$. There are two outcomes $Z = \{z, z'\}$, where z is the outcome of not forming a partnership and z' is the outcome of forming a partnership. Players' payoffs from not forming a partnership are always 0. Players' payoffs from forming a partnership, $(u_1(z', \theta), u_2(z', \theta))$, are dependent on their types $\theta = (\theta_1, \theta_2)$ as follows:

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$1, -2$
θ_1^2	$-2, -2$	$-2, -2$

This configuration means that, for instance, players 1 and 2 receive payoffs of 1 and -2, respectively, from the partnership if their types are $\theta = (\theta_1^1, \theta_2^2)$. Assume that each player can unilaterally choose not to form a partnership, but both parties must agree in order to form a partnership. Consider a playout that assigns the outcome z to every state, $\pi \equiv z$, and a belief system β, in which $\beta_n(N, z, z', \theta_n)$ assigns equal probabilities to the opponent's two types in his perceived scenario of deviation (N, z, z', θ_n^1) and

[7] See Osborne and Rubinstein (1994) for a textbook treatment of common knowledge and self-evident events.

(N, z, z', θ_n^2). Therefore, player 1 of type θ_1^1 prefers the deviation, i.e.,

$$\mathbf{E}_{\beta_1(N,z,z',\theta_1^1)}(u_1(z',\cdot)) = 1 > 0 = \mathbf{E}_{\beta_1(N,z,z',\theta_1^1)}(u_1(z,\cdot)),$$

but player 1 of type θ_1^2 and player 2 of both types prefer not to deviate:

$$\begin{aligned}
\mathbf{E}_{\beta_1(N,z,z',\theta_1^2)}(u_1(z',\cdot)) &= -2 < 0 = \mathbf{E}_{\beta_1(N,z,z',\theta_1^2)}(u_1(z,\cdot)),\\
\mathbf{E}_{\beta_2(N,z,z',\theta_2^1)}(u_2(z',\cdot)) &= -\frac{1}{2} < 0 = \mathbf{E}_{\beta_2(N,z,z',\theta_2^1)}(u_2(z,\cdot)),\\
\mathbf{E}_{\beta_2(N,z,z',\theta_2^2)}(u_2(z',\cdot)) &= -2 < 0 = \mathbf{E}_{\beta_2(N,z,z',\theta_2^2)}(u_2(z,\cdot)).
\end{aligned}$$

Therefore, (N, z, z') is not a viable deviation in any state θ. However, player 1 prefers the deviation if and only if his type is θ_1^1, i.e., $D_1^\beta(N, z, z') = \{\theta_1^1\}$. This incentive is understood by player 2 of both types and the relevant belief for player 2's decision of joining the coalitional deviation (N, z, z') should condition on this fact. That is, player 2 should assign probability 1 to player 1 being θ_1^1, instead of equal probability to θ_1^1 and θ_1^2, in any of his two perceived scenario of deviation, which is captured by weak consistency. Given this belief, player 2 of θ_2^1 will join the deviation (N, z, z') with player 1 of θ_1^1. Therefore, weak consistency therefore implies that $(N, z, z', \theta_1^1, \theta_2^1)$ is a viable scenario of deviation. It can be shown that the only playout that is part of a weakly consistent and stable assessment is the following:

	θ_2^1	θ_2^2
θ_1^1	z'	z
θ_1^2	z	z

So weak consistency identifies the the more intuitive outcomes in this game. This game is still special. In more general games, the updating of the prior $\beta_n^0(\cdot|\cdot)$ will play a more salient role. □

5.2 Strong Consistency

Weak consistency is a restriction on how each player's beliefs should align with the incentives faced by their opponents (which are determined by their beliefs). We shall show that it suffices to make strong predictions in applications. But it leaves open the possibility of making joint restrictions on all players' beliefs and incentives. The following example demonstrates the possibility in the simplest case.

Example 2. Consider the same partnership game as in Example 1, except that the

payoffs from forming a partnership, z', are as follows:

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$-2, -2$
θ_1^2	$-2, -2$	$-2, -2$

That is, forming a partnership is mutually beneficial if and only if players' types are (θ_1^1, θ_2^1). We will call θ_1^1 and θ_2^1 the "cooperative types." Consider a playout $\pi \equiv z$ and a belief system, in which $\beta_n(N, z, z', \theta_n)$ assigns equal probabilities to the opponent's two types. Then, for each $n \in N$ and $\theta_n \in \Theta_n$,

$$\mathbf{E}_{\beta_n(N,z,z',\theta_n)}(u_n(z', \cdot)) < \mathbf{E}_{\beta_n(N,z,z',\theta_n)}(u_n(z, \cdot)).$$

Hence, in no state will player n benefit from the deviation, i.e., $D_n^\beta(N, z, z') = \emptyset$. It follows that (π, β) is stable and weakly consistent.

However, it is quite intuitive that in this common-interest game, the two cooperative-type players can form a partnership in (θ_1^1, θ_2^1). For instance, player 1 of type θ_1^1 can make the following statement to player 2: "I am θ_1^1, and if you are θ_2^1, let's form a partnership. You should know that if you are θ_2^1, I benefit from the partnership if and only if my type is θ_1^1, so you should trust that I am θ_1^1. My question is whether you are θ_2^1." Similarly, player 2 of type θ_2^1 can make the following mirroring statement to player 1: "I am θ_2^1, and if you are θ_1^1, let's form a partnership. You should know that if you are θ_1^1, I benefit from the partnership if and only if my type is θ_2^1, so you should trust that I am θ_2^1. My question is whether you are θ_1^1." The two statements are "mutually reassuring" in the sense that conditional on that the opponent is the cooperative type, a player benefits from carrying out the stated plan if and only if he himself is the cooperative type, which reassures the cooperative opponent. The existence of such mutual reassurance means that it is plausible that the two players collectively deviate from z to z' in (θ_1^1, θ_2^1).

We can deduce that the only stable playout $\pi : \Theta \to Z$ that survives this reasoning is such that the outcome z' is obtained if and only if the state is (θ_1^1, θ_2^1). □

This example is reminiscent of trading under heterogeneous beliefs (Levin, 2003) and the blocking condition of credible core (Dutta and Vohra, 2005).[8] We need to introduce

[8]The no-trade theorem, as formulated by authors such as Milgrom and Stokey (1982) and Rubinstein and Wolinsky (1990), provides conditions under which no trade will occur under arbitrary *given* information structure. In contrast, the mutual reassurance in this example leads to the creation of new information, i.e., the revelation of (θ_1^1, θ_2^1).

new ideas to accommodate more complex applications, as shown in the example below.

Example 3. Consider a game with two outcomes $Z = \{z, z'\}$ between two players $N = \{1, 2\}$, where each player has two types and there is a uniform common prior $\beta^0 \in \Delta(\Theta)$. Players' payoffs from z are 0. The payoffs from z', $(u_1(z', \cdot), u_2(z', \cdot))$, depend on the state as follows:

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$-2, 1$
θ_1^2	$1, 1$	$3, -2$

Consider the playout $\pi : \Theta \to Z$ given by

	θ_2^1	θ_2^2
θ_1^1	z	z
θ_1^2	z'	z

In state (θ_1^2, θ_2^1), z' is the outcome, so the two players do not want to deviate to z. At the outcome z, the relevant payoff matrix is

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$-2, 1$
θ_1^2		$3, -2$

In state (θ_1^1, θ_2^1), the two players have a common interest to move from z to z', but knowing player 1's type is θ_1^1, player 2's type θ_2^2 would also participate in this deviation, deterring θ_1^1 from participating in the first place. This seems to unravel the coalitional deviation. However, the following "mutually reassuring" statements can sustain the common interest deviation from z to z' in (θ_1^1, θ_2^1) : player 1 states that he would join the deviation regardless of his types and asks player 2 to join only when player 2's type is θ_2^1; player 2 states that he will join the deviation only when his type is θ_2^1 and asks player 1 to join no matter what.

Let's see how mutual reassurance is created. Believing that player 1 will join the deviation regardless of his types, player 2 indeed finds it beneficial to join only when his type is θ_2^1. Believing player 2's claim that only θ_2^1 will join the deviation, player 1 finds it beneficial to join the deviation if his type is θ_1^1. How about θ_1^2? When player 1 is θ_1^2, he believes that player 2 is θ_2^2. However, according to player 2's claim, θ_2^2 will not join the deviation and, more importantly, will not find it beneficial to join the deviation

if both types of player 1 join the deviation. However, if player 2 trembles (i.e., join the deviation even if type θ_2^2 does not benefit from the deviation), then θ_1^2 indeed benefits from joining the deviation. Hence, player 1's claim of joining the deviation regardless of his types is justified.

The existence of mutual reassurance depends on payoff functions. Suppose the following payoff matrix associated with z' is instead given by

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$-2, 1$
θ_1^2	$1, 1$	$-3, -2$

Then the reasoning above breaks down and it will be impossible for the two players to deviate to z' from z in (θ_1^1, θ_2^1). □

To formulate the ideas we have proposed, we start with beliefs.

Definition 7. *Given player n's prior belief $\beta_n^0 \in \Delta(\Theta)$, a version of the conditional probability $\beta_n^0(\cdot|\cdot)$ is a* ***conscious conditional probability*** *if for any perceived scenario of deviation $(S, z, z', \theta_n) \in \Sigma_{\pi,n}$ and any $F_{-n} \subset \Theta_{-n}$, the following holds:*

(i) $\beta_n^0(\{\theta_n\} \times \Theta_{-n} | (\{\theta_n\} \times F_{-n}) \cap \pi^{-1}(z)) = 1$;

(ii) $\beta_n^0(\pi^{-1}(z) | (\{\theta_n\} \times F_{-n}) \cap \pi^{-1}(z)) = 1$.

The two properties of conscious conditional probability are consistent with the "self-recognition" and "knowledge of playout" properties of a belief system in Definition 2. Therefore, regardless of how he updates his belief, player n knows his own type θ_n and the outcome z. It allows for conditioning on zero probability events. Player n will attribute any surprises—we say F_{-n} is a surprise if $(\{\theta_n\} \times F_{-n}) \cap \pi^{-1}(z)$ is an empty set—to mistakes or trembles of his opponents. For instance, in Example 3, if $\theta_1 = \theta_1^2$ and $F_2 = \{\theta_2^1\}$, we have $(\{\theta_1^2\} \times F_2) \cap \pi^{-1}(z) = \{(\theta_1^2, \theta_2^1)\} \cap \pi^{-1}(z) = \emptyset$. Therefore, if only player 2 of type θ_2^1 is expected to join the deviation, player 1 of type θ_1^2 would be surprised but he rationalizes this surprise by assuming that player 2's type is actually θ_2^2 and this type trembles (i.e., joins the deviation by mistake when not supposed to). We do not need to explicitly formulate trembles, but obviously it can be done.[9]

We are ready to formalize mutually reassuring sets of types that support a deviation.

[9] See, e.g., Kohlberg and Reny (1997) for an analogous discussion.

Definition 8. *A collection of sets* $\{D_n\}_{n\in S}$*, where* $D_n \subset \pi_n^{-1}(z)$*, is* ***mutually reassuring*** *for a deviation* $(S, z, z') \in \Delta_\pi$ *if*

(i) $(D_S \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset$*, where* $D_S = \prod_{n\in S} D_n$*, and*

(ii) *for all* $n \in S$*, we have*

$$D_n = \left\{ \theta_n \in \pi_n^{-1}(z) : \begin{array}{l} \mathbf{E}_n^0(u_n(z', \cdot)|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z)) \\ > \mathbf{E}_n^0(u_n(z, \cdot)|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z)) \end{array} \right\}, \tag{5.3}$$

where the conditional expectation $\mathbf{E}_n^0(\cdot|\cdot)$ *is defined with respect to a version of conscious conditional probability* $\beta_n^0(\cdot|\cdot)$*.*

Mutually reassuring sets $\{D_n\}_{n\in S}$ for a deviation form a fixed point of (5.3). This is where Tarski's fixed-point theorem comes into play in applications. Let us break down the mathematical expression that defines mutual reassurance. Condition (i) says that the type profiles of the coalition S do not contradict with the outcome z. Condition (ii) says that player n's types that prefer z' to z, conditional on his opponents' types are in $D_{S\setminus\{n\}}$, are exactly D_n. The expected payoff is computed as follows. The belief of player n, who has type θ_n, is updated from his prior β_n^0 conditional on the outcome z and the types of other players from the coalition being in $D_{S\setminus\{n\}}$. This updated belief, $\beta_n^0(\cdot|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z))$, which is a conscious conditional probability that takes into account the possibility of surprises, leads to an expected payoff $\mathbf{E}_n^0(u_n(\cdot, \cdot)|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z))$. In Example 2, $D_1 = \{\theta_1^1\}$ and $D_2 = \{\theta_2^1\}$ are mutually reassuring; in Example 3, $D_1 = \{\theta_1^1, \theta_1^2\}$ and $D_2 = \{\theta_2^1\}$ are mutually reassuring.

Remark 3. Identifying and carrying out mutually reassured deviations can be a demanding task for real-world players. But this is a good assumption to make for theory building, which is not very different from the leap of faith required for complete information solution concepts (see the discussion following Definition 1). □

Ultimately, we need to find a way to incorporate the idea of mutual reassurance in an assessment (π, β). This culminates in the following definition.

Definition 9. *An assessment* (π, β) *is* ***strongly consistent*** *if the following two conditions are satisfied:*

(i) *it is weakly consistent, and*

(ii) *a deviation* $(S, z, z') \in \Delta_\pi$ *is viable, i.e.,* $(D_S^\beta(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset$*, if there exist mutually reassuring sets* $\{D_n\}_{n \in S}$ *for the deviation.*

The conceptual difference between weak consistency and strong consistency is subtler than one might expect, although their implications very different as demonstrated by Examples 1–3.

Remark 4. Weak consistency is defined in terms of deviating sets. Deviating sets, if non-empty, define a common knowledge event of mutual gains from deviation, where the event is derived from the given belief system β and hence no new information is created that is not already captured by β. Together with stability, weak consistency requires that there be no common knowledge of gains from deviation. In contrast, strong consistency is defined in terms of mutually reassuring sets. Mutually reassuring sets, if they are not already common knowledge according to β, will create new information that refines the belief system β. Together with stability, strong consistency means that no such new information can be created. □

Weak and strong consistencies differ in their exact connection with mutually reassuring sets. The following result makes it precise. Its proof is in Appendix A.1.

Theorem 2. *Relationship between weak and strong consistency.*

(i) *If* (π, β) *is weakly consistent, then a deviation* $(S, z, z') \in \Delta_\pi$ *is viable* ***only if*** *there exist mutually reassuring sets* $\{D_n\}_{n \in S}$ *for the deviation.*

(ii) *If* (π, β) *is strongly consistent, then a deviation* $(S, z, z') \in \Delta_\pi$ *is viable* ***if and only if*** *there exist mutually reassuring sets* $\{D_n\}_{n \in S}$ *for the deviation.*

We give conditions for the existence of stable assessments that are weakly or strongly consistent. The proof is long and constructive and has been relegated to Appendix A.2.

Theorem 3. *If* $\pi(\theta)$ *is a stable outcome of* Γ^θ *for each* $\theta \in \Theta$*, then there exists a belief system* β *such that* (π, β) *is weakly consistent and stable. If, in addition,* π *is one-to-one, then* (π, β) *is strongly consistent and stable.*

The classic reference for existence under complete information is Shapley (1967). Dubey and Shapley (1984) identify a general and useful class of totally balanced games. Perhaps unexpectedly, a playout π that leads to state-by-state stable outcomes is not guaranteed to be part of an assessment that is strongly consistent and stable.

Example 4. Consider a two-player game where $N = \{1,2\}$, $\Theta = \{\theta_1^1, \theta_1^2\} \times \{\theta_2^1, \theta_2^2\}$, $Z = \{z, z'\}$, and $\mathbb{S} = \{N, \{1\}, \{2\}\}$. The prior belief over Θ is uniform. Payoffs from an outside option z are always 0. Payoffs from the partnership z', $(u_1(z,\cdot), u_2(z,\cdot))$, depend on the state as follows:

	θ_2^1	θ_2^2
θ_1^1	$+3, -1$	$-1, +3$
θ_1^2	$-1, +3$	$+3, -1$

Assume that players can unilaterally take the outside option but have to reach a consensus for a partnership: $d(N,\cdot) = Z$ and $d(\{n\}, z') = d(\{n\}, z) = \{z\}$ for $n \in N$. The unique state-by-state complete-information stable playout is $\pi \equiv z$. Note that Θ_1 and Θ_2 are mutually reassuring because $\beta_0(\cdot|\Theta_n)$ is uniform and all types in Θ_{-n} prefer z' (with an expected payoff of 1) to z. Therefore, there does not exist β such that (π, β) is strongly consistent and stable. □

5.3 Variants of Strong Consistency

Strong consistency has some useful variants.

5.3.1 Strong Consistency with Certainty

Definition 10. *A collection of sets $\{D_n\}_{n\in S}$, where $D_n \subset \pi_n^{-1}(z)$, is **mutually reassuring with certainty** for a deviation $(S, z, z') \in \Delta_\pi$ if*

(i) $(D_S \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset$, *and*

(ii) *for all $n \in S$,*

$$D_n = \left\{ \theta_n \in \pi_n^{-1}(z) : \begin{array}{r} \text{(a) } (\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset, \\ \text{(b) } \mathbf{E}_n^0(u_n(z',\cdot)|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z)) \\ > \mathbf{E}_n^0(u_n(z,\cdot)|(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z)) \end{array} \right\}, \quad (5.4)$$

where $\mathbf{E}_n^0(\cdot|\cdot)$ is defined with respect to the usual conditional probability $\beta_n^0(\cdot|\cdot)$.

In contrast with Definition 8, mutual reassurance with certainty requires that D_n only contain types that do not contradict with their opponents' types being in $D_{S\setminus\{n\}}$ and that the outcome is z; that is, player n is certain that his opponents do not tremble, and hence $\beta_n^0(\cdot|\cdot)$ is well defined. In Example 2, $\{\theta_1^1\}$ and $\{\theta_2^1\}$ are mutually reassuring with certainty. In Example 3, mutually reassuring sets with certainty do not exist.

Definition 11. *An assessment* (π, β) *is* ***strongly consistent with certainty*** *if the following two conditions are satisfied:*

(i) *it is weakly consistent, and*

(ii) *a deviation* $(S, z, z') \in \Delta_\pi$ *is viable if there exist* $\{D_n\}_{n\in S}$ *that are mutually reassuring with certainty for the deviation.*

The connection between strong consistency and strong consistency with certainty warrants an investigation. The key is belief independence. In non-cooperative games, we often assume that types remain independent under posterior beliefs after any history in games with observable actions if they are independent under the prior belief. Analogous properties can be defined for cooperative games.

Definition 12. (i) *A playout* π *has* ***type independence*** *if for all* $\theta \in \pi^{-1}(z)$, $z \in \pi(\Theta)$, *and* $n \in N$,

$$\beta_n^0(\theta|\pi^{-1}(z)) = \prod_{m\in N} \beta_n^0(\theta_m|\pi^{-1}(z)).$$

(ii) *An assessment* (π, β) *has* ***belief independence*** *if for all* $(S, z, z') \in \Delta_\pi$, $n \in S$, $\theta_n, \theta_n' \in \pi_n^{-1}(z)$, *and* $\theta_{-n} \in \Theta_{-n}$,

$$\beta_n(S, z, z', \theta_n)(\theta_{-n}) = \beta_n(S, z, z', \theta_n')(\theta_{-n}) = \prod_{m\neq n} \beta_n(S, z, z', \theta_n)(\theta_m). \tag{5.5}$$

Type independence in (i) invokes the prior belief β_n^0. Belief independence in (ii) captures the idea that player n in each of his perceived scenarios of deviation believes that his opponents' types are independent, and this belief is independent of his own type. In a sense belief independence is stronger than type independence because it also applies to deviations. This intuition is confirmed in Lemma 1.

Lemma 1. *Suppose* $(\{n\}, z, z) \in \Delta_\pi$ *for all* $n \in N$ *and* $z \in \pi(\Theta)$. *Then if a weakly consistent assessment* (π, β) *has belief independence, the playout* π *has type independence.*

The assumption of $(\{n\}, z, z) \in \Delta_\pi$ is vacuous and does not affect any concept previously defined.

Theorem 4. *Suppose* π *has type independence. Then the following holds.*

(i) $\{D_n\}_{n\in S}$ *are mutually reassuring for* $(S, z, z') \in \Delta_\pi$ *if and only if they are mutually reassuring with certainty.*

(ii) (π, β) *is strongly consistent if and only if it is strongly consistent with certainty.*

5.3.2 Strong Consistency with Correlation

Mutual reassurance and strong consistency are building blocks for even stronger restrictions, which are of conceptual interest. For instance, we can have correlated coalitional deviations, where individuals within each sub-coalition mutually reassure themselves and sub-coalitions mutually reassure one another. Due to space constraints, we skip the elaboration of the idea here.

6 Economic Application 1: Matching

Building theories without applications is like telling a superhero story without villains. We invite all readers, sympathetic or not, to explore further. To begin, we study applications where the solution concept is undisputed for the complete-information setting, while the understanding of the incomplete-information setting remains limited.

For matching games described in Section 3, Theorem 1 implies the following.

Corollary 1. *Suppose that the matching game has essentially private values. Then* (π, β) *is a stable assessment if and only if for each* $\theta \in \Theta$, $\pi(\theta)$ *is a stable matching outcome for the complete information game* Γ^θ.

A matching game has **one-sided incomplete information** if either Θ_i is a singleton for all $i \in I$, or Θ_j is a singleton for all $j \in J$. The following result spells out a conceptual difference between one-sided and two-sided incomplete information.

Theorem 5. *Suppose the matching game has one-sided incomplete information, and* $u_n(z, \theta)$ *depends on* θ *only through* θ_{z_n}. *Then an assessment* (π, β) *is weakly consistent if and only if it is strongly consistent.*

The conclusion is quite interesting, because Examples 2–4 show that weak consistency and strong consistency have very different implications if there is two-sided incomplete information. Our later results will build on this conceptual distinction. Both Corollary 1 and Theorem 5 hold with or without transfers and with or without outcome externality (i.e., payoffs depend on the matching outcome z of the whole market). Theorem 5 relies on the absence of information externality of θ_{-z_n}.

For the rest of this application, we consider a matching game with transfers. A matching outcome is $z = (\mu, \tau)$. Let M_n be the set of feasible matching partners for

player n; thus, $M_n = I \cup \{n\}$ if $n \in J$ and $M_n = J \cup \{n\}$ if $n \in I$. Players have **quasi-linear utility functions**: for all $n \in N$, there exists $v_n : M_n \times \Theta \to \mathbb{R}$ such that

$$u_n(\mu, \tau, \theta) = v_n(\mu_n, \theta) + \tau_n.$$

We shall allow for information externality. We also assume that players share a **common prior** $\beta^0 \in \Delta(\Theta)$.

6.1 Efficiency: An Outside Observer's Perspective

Starting with the prior β^0, knowing the playout π, and observing matching data z, an outside observer who doesn't possess insiders' private information will have a posterior distribution $\beta^0(\cdot|\pi^{-1}(z)) \in \Delta(\Theta)$. The expected surplus associated with this matching outcome computed from this posterior is $\sum_{n\in N} \mathbf{E}^0(u_n(z,\cdot)|\pi^{-1}(z))$.

We ask the following question: can the observer recommend a rearrangement of the matching to improve the expected surplus? That is to ask whether the following hold:

$$z \in \operatorname*{argmax}_{z' \in Z} \sum_{n\in N} \mathbf{E}^0(u_n(z',\cdot)|\pi^{-1}(z)). \tag{6.1}$$

If (6.1) holds for *all* $z \in \pi(\Theta)$, we say π is **Bayesian efficient**. If π is efficient in this sense, the outsider observer will not be able to make surplus-improving recommendations based on observed matching data alone, without changing the information structure. This criterion of the outside observer assessed efficiency has useful implications, as explained in the introduction.[10]

When information is complete, a duality exists between efficiency and stability. However, this duality does not always apply under incomplete information, even when considering this restricted notion of efficiency. Therefore, our objective is to identify general and economically meaningful conditions that preserve this duality.

The evaluation of expected surplus involves the knowledge of playout π and the incomplete-information game (in particular, payoff functions $(u_n)_{n\in N}$ and the common prior β^0). This is perhaps too much information for an outside observer to acquire. We are after a robust efficiency result—the planner needs to know neither the specific playout nor the exact game.

[10] This notion is adopted in Liu (2020) for the special case of one-sided incomplete information. It is different from efficiency criteria analyzed in Holmström and Myerson (1983) and Forges (1994), because it does not condition on each player's types although it conditions on the observed outcome.

Definition 13. *A matching game has* ***one-sided interdependence*** *if either there is no restriction on v_j but there exist functions $A_i : \Theta \to \mathbb{R}$ and $A'_i : M_i \to \mathbb{R}$ such that*

$$v_i(j, \theta) = A_i(\theta) + A'_i(j) \text{ for all } i \in I,\ j \in M_i,\ \text{and } \theta \in \Theta,$$

or symmetrically, there is no restriction on v_i but there exist functions $B_j : \Theta \to \mathbb{R}$ and $B'_j : M_j \to \mathbb{R}$ such that

$$v_j(i, \theta) = B_j(\theta) + B'_j(i) \text{ for all } j \in J,\ i \in M_j,\ \text{and } \theta \in \Theta.$$

Obviously, one-sided interdependence is different from one-sided incomplete information. The following are two special cases of one-sided interdependence.

$$v_i(j, \theta) = A_i(\theta_i) \text{ with no restriction on } v_j;$$
$$v_j(i, \theta) = B_j(\theta_j) \text{ with no restriction on } v_i.$$

One-sided interdependence appears natural in applications: workers' costs from work depend only on their types but their outputs depend on both theirs and firms' types, or producers' costs are their private information but the customers' payoffs depend on the private information of both sides. One-sided interdependence and weak consistency have strong implications.

Theorem 6. *If the matching game has one-sided interdependence, then the playout π of any weakly consistent and stable assessment (π, β) is Bayesian efficient.*

The proof is relegated to Appendix A.6. Under one-sided interdependence, the deviating set for one of the deviating players is either empty or the whole set of types. In this case, Bayes' rule under weak consistency imposes a very strong restriction on stability, which restores efficiency. We shall introduce a general class of matching games with what we call "comonotonic differences." To deal with the bigger class, we strengthen the weak consistency requirement to strong consistency.

6.2 Comonotonic Differences

Let X_1, X_2, and X_3 be finite sets. Consider two real-valued functions $f, g : X_1 \times X_2 \times X_3 \to \mathbb{R}$ and a weight function $w : X_3 \to \mathbb{R}_+$ on X_3. Define $f_w, g_w : X_1 \times X_2 \to \mathbb{R}$ as

$$f_w(\cdot) = \sum_{x_3 \in X_3} f(\cdot, x_3) w(x_3) \text{ and } g_w(\cdot) = \sum_{x_3 \in X_3} g(\cdot, x_3) w(x_3).$$

Definition 14. *We say that f and g are **comonotonic** on X_1 and X_2 if for any weight function $w : X_3 \to \mathbb{R}_+$, there exist total orders $\geq_1^w$ on X_1 and $\geq_2^w$ on X_2 such that both f_w and g_w are non-decreasing on X_1 and X_2.*

Comonotonicity on X_1 and X_2 is different from (co)monotonicity on $X_1 \times X_2$. It is stronger than the combination of monotonicity on X_1 for each $x_2 \in X_2$ and monotonicity on X_2 for each $x_1 \in X_1$.

Definition 15. *A matching game has **comonotonic differences** if $v_i(j,\theta) - v_i(j',\theta)$ and $v_j(i,\theta) - v_j(i',\theta)$ are comonotonic on Θ_i and Θ_j for any two pairs $(i,j) \in I \times J$ and $(i',j') \in M_j \times M_i$.*

The condition of comonotonic differences is new to the literature, yet it is both intuitive and useful for incomplete-information matching. The motivation is as follows. For any putative matching, consider a potential "blocking pair" i and j whose partners are $j' \neq j$ and $i' \neq i$, respectively. Worker i's ex post gain from the deviation is $v_i(j,\theta) - v_i(j',\theta)$ and firm j's is $v_j(i,\theta) - v_j(i',\theta)$. Comonotonic differences ensure that there are total orders on Θ_i and Θ_j for any two pairs (i,j,i',j') according to which the deviating incentives of i and j respond to their private information in the same direction. It is worthwhile to emphasize that the total orders on Θ_i and Θ_j are *not* a priori fixed: they can change with (i,j,i',j') and beliefs over Θ_{-ij}. As such, comonotonic differences does not rely on the monotonicity of matching values.

Some special cases of comonotonic differences are of interest in their own right.

- **Complete-Information Games.** Comonotonic differences places *no* restriction on complete-information matching games. So it is not an extension of any known condition for complete information matching problems.

- **One-sided Interdependence.** To verify comonotonic differences, consider the first case where there is no restriction on v_j. Then

  $$v_i(j,\theta) - v_i(j',\theta) = A_i'(j) - A_i'(j')$$

  does not depend on θ_i and θ_j. Therefore, $v_i(j,\theta) - v_i(j',\theta)$ and $v_j(i,\theta) - v_j(i',\theta)$ are comonotonic on Θ_i and Θ_j.

- **Separable Values.** A matching game has *separable values* if

$$v_i(j,\theta) = A_i(\theta) + A_i'(j,\theta_{-i}) \text{ for all } i \in I,\ j \in M_i, \text{ and } \theta \in \Theta,$$
$$v_j(i,\theta) = B_j(\theta) + B_j'(i,\theta_{-j}) \text{ for all } j \in J,\ i \in M_j, \text{ and } \theta \in \Theta,$$

  where $A_i, B_j : \Theta \to \mathbb{R}$, $A_i'(j,\cdot) : \Theta_{-i} \to \mathbb{R}$ and $B_j'(i,\cdot) : \Theta_{-j} \to \mathbb{R}$. Matching games with separable values have been useful for empirical analysis.

  To see that the condition of separable values implies comonotonic differences, observe that $v_i(j,\theta) - v_i(j',\theta) = A_i(j,\theta_{-i}) - A_i(j',\theta_{-i})$, which is independent of θ_i, and $v_j(i,\theta) - v_j(i',\theta) = B_j(i,\theta_{-j}) - B_j(i',\theta_{-j})$, which is independent of θ_j. Therefore, $v_i(j,\theta) - v_i(j',\theta)$ and $v_j(i,\theta) - v_j(i',\theta)$ are comonotonic on Θ_i and Θ_j.

- **Common Values.** Consider a two-player coordination game with incomplete information: $I = \{i\}$ and $J = \{j\}$. Also $v_i(j,\cdot) = v_j(i,\cdot)$ and $v_i(i,\cdot) = v_j(j,\cdot) \equiv 0$. The game has comonotonic differences. The uniform sharing rule considered by Dizdar and Moldovanu (2016), where $v_i(j,\cdot) = \lambda v_j(i,\cdot)$, $\lambda > 0$ is a scalar, also satisfies comonotonic differences.

The condition of omonotonic differences connects efficiency and stability.

Theorem 7. *If the game has comonotonic differences, then the playout π of any strongly consistent and stable assessment (π, β) with type independence is Bayesian efficient.*

We would like to emphasize that although strong consistency and type independence implies strong consistency with certainty (Theorem 4), the latter is not sufficient to ensure efficiency and counterexamples are not difficult to concoct. Independence is needed to wash out the information externality of the private information of players not involved in the deviation (but their payoff externality is not ruled out). The proof in Appendix A.8 employs two tools: strong duality and Tarski's fixed point theorem.

First, by the dual program of surplus maximization (6.1), a failure of efficiency is translated into what we call "auxiliary deviations" that do not condition on players' private information or each other's incentives to deviation (Lemma 3 and Lemma 4 in Appendix A.5). In a nutshell, an auxiliary deviation from z involves a pair of players (i,j) and a transfer such that the following hold:

$$\begin{aligned} &\mathbf{E}^0(i\text{'s gain from deviation}|\pi^{-1}(z)) > 0, \\ &\mathbf{E}^0(j\text{'s gain from deviation}|\pi^{-1}(z)) > 0. \end{aligned} \tag{6.2}$$

Under complete information, the existence of an auxiliary deviation implies that z is not a stable outcome, as is well-known. With incomplete information, the auxiliary deviation only reflects an outside observer's perspective, which is not how players inside the game look at the deviation. It follows from Theorem 2 that, under strong consistency, a viable deviation for these players must be supported by non-empty mutually reassuring sets D_i and D_j that satisfy the following fixed-point property:

$$\begin{aligned} D_i &= \{\theta_i : \mathbf{E}^0(i\text{'s gain from deviation}|\pi^{-1}(z), \theta_i, D_j) > 0\}, \\ D_j &= \{\theta_j : \mathbf{E}^0(j\text{'s gain from deviation}|\pi^{-1}(z), \theta_j, D_i) > 0\}. \end{aligned} \tag{6.3}$$

Second, to go from an auxiliary deviation of an outsider's perspective (6.2) to a viable deviation of insiders' perspective (6.3), we invoke comonotonic differences. The existence of non-empty mutually reassuring sets is an application of Tarski's fixed point theorem (Lemma 5 in Appendix A.7). Example 2 shows that the condition of strong consistency cannot be relaxed in the result.

7 Economic Application 2: Networks

The basic setup is described in Section 3.[11] We further assume that each player's type set is linearly ordered. Player n's payoff from a network z is

$$u_n(z, \theta) = \sum_{m \in z_n} v_{nm}(\theta_n, \theta_m) - c_n(|z_n|, \theta_n),$$

where $v_{nm} : \Theta_n \times \Theta_m \to \mathbb{R}$ specifies the value of the link (n, m) to player n, and $c_n : \mathbb{N} \times \Theta_n \to \mathbb{R}$ specifies player n's cost of maintaining his links.[12]

We introduce two more modifications that are easily incorporated into the main theoretical framework. First, in a putative network, if $m \in N_n$ (i.e., m is connected to player n), then player n observes m's type θ_m. So effectively, player n's information type is θ_{z_n}. We do not assume that indirectly linked players, or disconnected players trying to form a new connection, have knowledge of each other's types.

Second, prior to forming a new link with player m in an existing network, player $n \neq m$ can take a costly action $s_n \in \mathbb{R}_+$ that is observable to m (see the introduction for applications of such signaling activities; the question of whether player n's action

[11] The cooperative model simplifies the game of network formation, which studies a given network's robustness to deviations; see, e.g., Dutta and Mutuswami (1997) for a rich strategic model.

[12] This payoff specification is not uncommon in the literature; see, e.g., Sadler (2022).

is observable to a third party is irrelevant for the solution concept, because it tests the network's vulnerability to any single deviation instead of a chain of deviations). Player n's cost function is $C_{nm} : \mathbb{R}_+ \times \Theta_n \times \Theta_m \to \mathbb{R}$, so player n's ex post signaling cost can depend on player m's private type, as well as player m's observable attributes, which are summarized by m. Importantly, the action is assumed to be a non-productive signaling device. We allow the ex post signaling cost to depend on both players' types for generality. Player m can also take a costly action s_m. Therefore, given any network z, a deviation $\delta = (\{m, n\}, z, z', s_m, s_n)$ involves a pair of players m and n, who take signaling actions s_m and s_n, respectively, and form a new link between them to obtain the network z'. Player n's *ex post* payoff from the deviation is $u_n(z', \theta) - C_{nm}(s_n, \theta_n, \theta_m)$, while his payoff from not deviating is $u_n(z, \theta)$. We shall make the following assumptions.

Assumption 1.

(i) *v_{nm} is strictly increasing in both θ_n and θ_m for all $n \neq m$;*

(ii) *$c_n(k+1, \cdot) - c_n(k, \cdot)$ is non-increasing in θ_n for all $k \geq 1$ (i.e., high types have weakly lower marginal cost of maintaining a link);*

(iii) *$C_{nm}(0, \cdot) \equiv 0$, $\lim_{s_n \to \infty} C_{nm}(s_n, \theta_n, \theta_m) = \infty$ for each θ_n and θ_m; C_{nm} is strictly increasing and continuous in s_n and non-increasing in θ_n (i.e., signaling is less expensive for high types, but the cost needn't be monotonic in the other player's types).*

We do not assume that the cost of maintaining links, c_n, is increasing or convex in the number of links, k. A special case of the signaling cost function is $C_{nm}(s_n, \theta_n, \theta_m) \equiv s_n$ so that signaling is purely money burning.

With the two modifications, a deviation is $\delta = (\{m, n\}, z, z', s_m, s_n)$, a scenario of deviation is $\sigma = (\delta, \theta)$, and player n's perception of the scenario is $\sigma = (\delta, \theta_{z_n})$.[13] Our goal is to understand the structure of stable networks under incomplete information.

Theorem 8. *If (π, β) is strongly consistent with certainty and stable, then, for any $\theta \in \Theta$, $z = \pi(\theta)$ is a stable outcome of the complete information game Γ^θ.*

Strong consistency and type independence imply strong consistency with certainty. But unlike Theorem 7, independence is not needed in Theorem 8. How can we reconcile

[13] Therefore, the definitions of stability and consistencies should treat player n's type as θ_{z_n}. Details are fleshed out in Appendix A.9.

the difference? Isn't that both matching and networks are concerned with pairwise stability? The difference is that in network formation problem, especially the stability concept of Jackson and Wolinsky (1996), players are not required to abandon their existing partners before forming a new link, and hence the types of other players will not play a role. This is clearly not the case in one-to-one matching, where deviating players must abandon their partners whose types are informationally relevant. Both concepts are reasonable. Incomplete information accentuates their subtle difference.

We should clarify the scope of the result. To understand networks, it is crucial to distinguish between stable network structures and network formation processes. Theorem 8 states that, regardless of the network formation process, we should expect the patterns of connections between nodes under incomplete information to resemble those under complete information. However, it is important to note that players may still be uncertain about the types of players they are not connected to, so uncertainty does not fully unravel. This offers insight into how incomplete-information networks function in the long run and the role of incomplete information, without relying on specific assumptions about network formation processes.

But costly mutual signaling indeed indirectly captures a consequential aspect of network formation processes. Signaling takes place when forming new connections, which, in our formulation, is a deviation from a stabilized network. Therefore, our modeling has achieved the separation of network structures and network formation. Theorem 8 shows that this modeling of network formation as a deviation scenario leads to stark implications for network structures. The approach of disentangling stabilized network structures and network formation should be explored further. For example, one useful direction is to measure the inefficiency of network formation by studying the minimal total signaling cost it must entail for an unstable network structure to stabilize.

In the proof, we show that as long as players m and n find it mutually profitable to form a link under the complete information of (θ_m^*, θ_n^*), they can take costly actions s_m and s_n, respectively, and mutually reassure a deviation in *some* state under incomplete information (which may not be the same as (θ_m^*, θ_n^*)). The literature of signaling games offers many ideas about equilibrium refinements, which could be useful to extend strengthen our notion of strong consistency, although most of these games have one-sided incomplete information. For cooperative analysis, the exact signaling action

is not important—what is relevant is whether there is a deviation that will destabilize z—we again test a solution against all possible deviations.

8 Conclusion

This paper presents a general approach to cooperative games with incomplete information and studies two of its applications that have been limited in progress due to their complexity. As reduced-form models of strategic interactions without making excessive ad hoc assumptions, cooperative games will have many more interesting applications, in which deviations can be bilateral or multilateral. The advantages of cooperative analysis over non-cooperative analysis come at a cost. Specifically, it does not model strategic behaviors, and due to the simplicity of cooperative models, unique predictions are often difficult to obtain. Consequently, a delicate balance must be struck between simplifying our models and obtaining meaningful results. We believe that concrete economic applications can provide valuable guidance.

We leave open the question of implementing solution concepts or, more broadly, a Nash program for incomplete-information games (see Okada (2012), Kamishiro, Vohra, and Serrano (2022) for seminal contributions).[14] One immediate idea is to require that playouts are incentive compatible. All of our results still hold with this additional condition except that establishing existence is not generally possible. However, this mechanism design approach is too restrictive, particularly in the context of observable outcomes, because playouts are intended to be stabilized allocations as a result of complex interactions that are not modeled; to complement the mechanism design approach, the outcome space needs to be augmented in some way to account for the processes leading to these outcomes. It should also be mentioned that the efficiency notion we consider in this paper is allocative efficiency, which does not rule out inefficient delays of these progresses.[15] Note that stability and familiar notions of implementability are known to be in conflict even in complete-information applications such as two-sided matching except in the special case where the preference of one side of the market is

[14]This exercise is quite subtle even under complete information; see, e.g., Gul (1989), Perry and Reny (1994), etc.

[15]Inefficient delays in frictional bargaining (due to incomplete information or non-negligible discounting) should not surprise us; see, e.g., Rubinstein and Wolinsky (1985).

ignored.

In this paper, we focus on the solution concepts of core and stability. The long history of cooperative game theory is marked by the development of a wealth of useful solution concepts for games involving complete information: values, bargaining solutions, the nucleolus, the von Neumann–Morgenstern stable sets, etc. Ray (2007) offers a systematic examination of ideas to coalition formation with complete information. Gul and Pesendorfer (2020) propose a new agenda for using Lindahl equilibrium and set-valued bargaining solutions in collective choice and market design problems. Our approach of formulating solution concepts can be used to extend these cooperative concepts to incomplete-information environments.

A Appendix

A.1 Proof of Theorem 2

Lemma 2. *Suppose (π, β) is weakly consistent and $\{D_n^\beta(S,z,z')\}_{n\in S}$ are deviating sets for a deviation $(S,z,z') \in \Delta_\pi$. If $(D_S^\beta(S,z,z')\times\Theta_{-S})\cap\pi^{-1}(z) \neq \emptyset$, then $\{D_n^\beta(S,z,z')\}_{n\in S}$ are mutually reassuring for the deviation (S,z,z').*

Proof. We only need to verify the fixed-point property (5.3) in the definition of mutually reassuring sets. Combining (5.1) and (5.2), we have

$$D_n^\beta(S,z,z') = \left\{\theta_n \in \pi_n^{-1}(z) : \begin{array}{l} \mathbf{E}_n^0(u_n(z',\cdot)|(\{\theta_n\}\times D_{S\setminus\{n\}}^\beta(S,z,z')\times\Theta_{-S})\cap\pi^{-1}(z)) \\ > \mathbf{E}_n^0(u_n(z,\cdot)|(\{\theta_n\}\times D_{S\setminus\{n\}}^\beta(S,z,z')\times\Theta_{-S})\cap\pi^{-1}(z)) \end{array}\right\} \quad \text{(A.1)}$$

where the conditional expectation is with respect to the belief in (5.2):

$$\beta_n(S,z,z',\theta_n)(\cdot) = \beta_n^0\left(\cdot|(\{\theta_n\}\times D_{S\setminus\{n\}}^\beta(S,z,z')\times\Theta_{-S})\cap\pi^{-1}(z)\right). \quad \text{(A.2)}$$

Notice that (A.2) is a conscious conditional probability because the belief system satisfies "self-recognition" and "knowledge of playout." Thus, $\{D_n^\beta(S,z,z')\}_{n\in S}$ satisfy (5.3). □

Proof of Theorem 2. (i) By Definition 3 and the definition of deviating sets, the deviation (S,z,z') is viable if and only if $(D_S^\beta(S,z,z')\times\Theta_{-S})\cap\pi^{-1}(z) \neq \emptyset$. It follows from Lemma 2 that $\{D_n^\beta(S,z,z')\}_{n\in S}$ are mutually reassuring for the deviation (S,z,z').

(ii) The "only if" part follows from (i) above. The "if" part follows from the definition of strong consistency. □

A.2 Proof of Theorem 3

Proof. The proof is constructive. Unfortunately, we cannot find a shorter and simpler argument. For each deviation $(S, z, z') \in \Delta_\pi$, we denote by $\pi^{-1}_{S\setminus\{n\}}(z)$ the projection of $\pi^{-1}(z)$ on $\Theta_{S\setminus\{n\}}$. The proofs proceed in four steps.

Step 1. *Defining the support of* $\beta_n(S, z, z', \theta_n)$.

For each $n \in S$ and $\theta_n \in \pi^{-1}(z)$, we shall define a Boolean function $a_{n,\theta_n} : \pi_n^{-1}(z) \to \{0, 1\}$ such that $a_{n,\theta_n}(\theta) = 1$ if and only if θ is in the support of $\beta_n(S, z, z', \theta_n)$.

Case 1. For any $n \in S$ and $\theta_n \in \pi_n^{-1}(z)$, if $u_n(z', \theta_n, \theta_{-n}) > u_n(z, \theta_n, \theta_{-n})$ for all θ_{-n} such that $(\theta_n, \theta_{-n}) \in \pi^{-1}(z)$, let $a_{n,\theta_n}(\theta_n, \theta_{-n}) = 1$. Let Θ_n^* be the set of all such θ_n's.

Case 2. If there exist $n \in S$ and $\bar\theta_{S\setminus\{n\}} = (\bar\theta_m)_{m\in S\setminus\{n\}} \in \pi^{-1}_{S\setminus\{n\}}(z)$ such that $u_m(z', \bar\theta_m, \theta_{-m}) > u_m(z, \bar\theta_m, \theta_{-m})$ for all $m \in S\setminus\{n\}$ and $(\bar\theta_m, \theta_{-m}) \in \pi^{-1}(z)$, then for all $\theta = (\theta_n, \bar\theta_{S\setminus\{n\}}, \theta_{-S}) \in \pi^{-1}(z)$, we have $u_n(z', \theta) \le u_n(z, \theta)$, because otherwise we would have (S, z, z') as a viable deviation for Γ^θ. For all such n and θ, we let $a_{n,\theta_n}(\theta) = 1$.

Case 3. For any $n \in S$ and $\theta_n \in \pi_n^{-1}(z)\setminus(\Theta_n^* \cup \Theta_n^{**})$, there exist θ_{-n} such that $\theta := (\theta_n, \theta_{-n}) \in \pi^{-1}(z)$ and $u_n(z', \theta) \le u_n(z, \theta)$, because otherwise we would have $\theta_n \in \Theta_n^*$. We let $a_{n,\theta_n}(\theta) = 1$.

Finally, for any $n \in S$ and $\theta = (\theta_n, \theta_{-n}) \in \pi^{-1}(z)$, let $a_{n,\theta_n}(\theta) = 0$ if $a_{n,\theta_n}(\theta)$ is not yet defined in Cases 1–3.

The following properties follow immediately from the definition of a_{n,θ_n}:

(i) for any $\theta_n \in \pi_n^{-1}(z)$, there exists $\theta = (\theta_n, \theta_{-n}) \in \pi^{-1}(z)$ such that $a_{n,\theta_n}(\theta) = 1$;

(ii) if $\theta_n \neq \theta'_n$, then $a_{n,\theta_n}(\theta'_n, \theta'_{-n}) = 0$ for any $(\theta'_n, \theta'_{-n}) \in \pi^{-1}(z)$;

(iii) if $a_{n,\theta_n}(\theta_n, \theta_{-n}) = 1$, then either $u_n(z', \theta_n, \theta_{-n}) \le u_n(z, \theta_n, \theta_{-n})$ or $\theta_n \in \Theta_n^*$, i.e.,

$$u_n(z', \theta_n, \theta_{-n}) > u_n(z, \theta_n, \theta'_{-n})$$

for all θ'_{-n} such that $(\theta_n, \theta'_{-n}) \in \pi^{-1}(z)$.

Step 2. *Defining* $\beta_n(S, z, z', \theta_n)$.

For any $n \in S$ and $(S, z, z', \theta_n) \in \Sigma_{\pi,n}$, we define

$$\beta_n(S, z, z', \theta_n)(\cdot) = \beta_n^0(\cdot \mid \{\theta : a_{n,\theta_n}(\theta) = 1\}). \tag{A.3}$$

By property (i) in Step 1, $\{\theta : a_{n,\theta_n}(\theta) = 1\} \neq \emptyset$, and $\beta_n(S, z, z', \theta_n)(\theta) > 0$ if and only if $a_{n,\theta_n}(\theta) = 1$. It follows from property (ii) in Step 1 that β so defined is a belief system.

Step 3. *Showing that* (π, β) *is weakly consistent.*

Consider $n \in S$ and $(S, z, z', \theta_n) \in \Sigma_{\pi,n}$. We consider the three cases: $\theta_n \in \Theta_n^*$, $\theta_n \in \Theta_n^{**}$, and $\theta_n \in \pi_n^{-1}(z) \backslash (\Theta_n^* \cup \Theta_n^{**})$.

Suppose that $\theta_n \in \Theta_n^*$. We claim that for any $\theta = (\theta_m)_{m \in N} \in \pi^{-1}(z)$ such that $a_{n,\theta_n}(\theta) = 1$, there exists $m \in S \backslash \{n\}$ such that for all $(\theta_m, \theta'_{-m}) \in \pi^{-1}(z)$, $a_{m,\theta_m}(\theta_m, \theta'_{-m}) = 1$ only if $u_m(z', \theta_m, \theta'_{-m}) \leq u_m(z, \theta_m, \theta'_{-m})$. It follows from the claim that

$$(\{\theta_n\} \times D^{\beta}_{S \backslash \{n\}}(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) = \emptyset$$

and hence, weak consistency imposes no additional restriction on $\beta_n(S, z, z', \theta_n)$. To prove the claim, suppose to the contrary that there exists $\theta = (\theta_m)_{m \in N}$ such that $a_{n,\theta_n}(\theta) = 1$, and for all $m \in S \backslash \{n\}$ there exists θ'_{-m} such that $a_{m,\theta_m}(\theta_m, \theta'_{-m}) = 1$ and $u_m(z', \theta_m, \theta'_{-m}) > u_m(z, \theta_m, \theta'_{-m})$. By the definition of a_{m,θ_m} (property (iii) in Step 1), $\theta_m \in \Theta_m^*$ for all $m \in S \backslash \{n\}$. Together with $\theta_n \in \Theta_n^*$, it implies that (S, z, z') is a viable deviation for Γ^θ, a contradiction.

Suppose that $\theta_n \in \Theta_n^{**}$. Then

$$(\{\theta_n\} \times D^{\beta}_{S \backslash \{n\}}(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) = \{\theta : a_{n,\theta_n}(\theta) = 1\}.$$

Weak consistency requires that $\beta_n(S, z, z', \theta_n)(\cdot) = \beta_n^0(\cdot | \{\theta : a_{n,\theta_n}(\theta) = 1\})$, which is satisfied because of (A.3).

Suppose $\theta_n \in \pi_n^{-1}(z) \backslash (\Theta_n^* \cup \Theta_n^{**})$. Then we claim that for each $\theta = (\theta_m)_{m \in N} \in \pi^{-1}(z)$ there is an $m \in S \backslash \{n\}$ such that

$$\mathbf{E}_{\beta_m(S,z,z',\theta_m)}(u_m(z', \cdot)) \leq \mathbf{E}_{\beta_m(S,z,z',\theta_m)}(u_m(z, \cdot)).$$

It follows from the claim that $(\{\theta_n\} \times D^{\beta}_{S \backslash \{n\}}(S, z, z') \times \Theta_{-S}) \cap \pi^{-1}(z) = \emptyset$ and, hence, weak consistency places no additional restriction on $\beta_n(S, z, z', \theta_n)$. To prove the claim, suppose to the contrary that there exists $\theta = (\theta_m)_{m \in N} \in \pi^{-1}(z)$ such that

$$\mathbf{E}_{\beta_m(S,z,z',\theta_m)}(u_m(z', \cdot)) > \mathbf{E}_{\beta_m(S,z,z',\theta_m)}(u_m(z, \cdot))$$

for all $m \in S \backslash \{n\}$. Then by the definition of a_{m,θ_m} (property (iii) in Step 1) and the definition of $\beta_m(S, z, z', \theta_m)$, we have $\theta_m \in \Theta_m^*$. Therefore, $\theta_n \in \Theta_n^* \cup \Theta_n^{**}$, a contradiction.

Step 4. *Showing that* (π, β) *is stable.*

Consider any scenario of deviation $(S, z, z', \theta) \in \Sigma_\pi$. If there exists $n \in S$ such that $\theta_n \notin \Theta_n^*$, then player n will not find the deviation profitable (by property (iii) in Step 1). If $\theta_n \in \Theta_n^*$ for all $n \in S$, then $\pi(\theta)$ is not a stable outcome for Γ^θ (again by property

(iii) in Step 1), a contradiction. Therefore, (S, z, z', θ) is not a viable scenario.

Step 5. *Strong consistency and strong consistency with certainty*

If π is one-to-one, then $\pi^{-1}(z)$ is a singleton and $\beta_n(S, z, z', \theta_n)$ assigns probability 1 to it. The only candidate mutually reassuring sets are $\{\pi_n^{-1}(z)\}_{n\in S}$. Therefore, strong consistency and strong consistency with certainty are always satisfied. □

A.3 Proof of Theorem 4

Proof of Lemma 1. If $S = \{n\}$, then $(D^{\beta}_{S\setminus\{n\}} \times \Theta_{-(S\setminus\{n\})}) \cap \pi^{-1}(z) = \pi^{-1}(z) \neq \emptyset$. Belief independence and weak consistency imply that, for any $(\{n\}, z, z, \theta_n) \in \Sigma_{\pi,n}$ and $m \neq n$,

$$\beta_n(\{n\}, z, z, \theta_n)(\theta_{-n}) = \beta_n^0(\theta_{-n}|(\{\theta_n\} \times \Theta_{-n}) \cap \pi^{-1}(z)) = \beta_n^0(\theta_{-n}|\pi^{-1}(z)), \quad \text{(A.4)}$$

$$\beta_n(\{n\}, z, z, \theta_n)(\theta_m) = \beta_n^0(\theta_m|(\{\theta_n\} \times \Theta_{-n}) \cap \pi^{-1}(z)) = \beta_n^0(\theta_m|\pi^{-1}(z)), \quad \text{(A.5)}$$

where the last inequalities of both (A.4) and (A.5) follow because $\beta_n(\{n\}, z, z, \theta_n)(\theta_{-n})$ is independent of $\theta_n \in \pi^{-1}(z)$. By belief independence, we have

$$\beta_n(\{n\}, z, z, \theta_n)(\theta_{-n}) = \prod_{m\neq n} \beta_n(\{n\}, z, z, \theta_n)(\theta_m). \quad \text{(A.6)}$$

From (A.4)–(A.6), we have $\beta_n^0(\theta_{-n}|\pi^{-1}(z)) = \prod_{m\neq n} \beta_n^0(\theta_m|\pi^{-1}(z))$. Therefore,

$$\beta_n^0(\theta|\pi^{-1}(z)) = \beta_n^0(\theta_{-n}|(\{\theta_n\} \times \Theta_{-n}) \cap \pi^{-1}(z))\beta_n^0(\theta_n|\pi^{-1}(z)) = \prod_{n\in N} \beta_n^0(\theta_n|\pi^{-1}(z)).$$

That is, π has type independence. □

Proof of Theorem 4. Since $\pi^{-1}(z)$ is a product space, then for any $\theta_n \in \pi_n^{-1}(z)$ and any $D_m \subset \pi_m^{-1}(z)$, then $(\{\theta_n\} \times D_{S\setminus\{n\}} \times \Theta_{-S}) \cap \pi^{-1}(z) \neq \emptyset$ and hence (a) in (5.4) is always satisfied and the conventional conditional probability $\beta_n^0(\cdot|\cdot)$ is a conscious conditional probability. Hence (5.3) and (5.4) are identical. The results then follow immediately. □

A.4 Proof of Theorem 5

Proof. Suppose $\Theta_j = \{\theta_j^*\}$ is a singleton for all $j \in J$. If (π, β) is weakly consistent but not strongly consistent, then there exists a deviation $(\{i, j\}, z, z')$, which is not viable, but there are mutually reassuring sets D_i and D_j for the deviation, where $D_j = \{\theta_j^*\}$ and

$$D_i = \{\theta_i \in \pi_i^{-1}(z) : u_i(z', \theta_i, \theta_j^*) > u_i(z, \theta_i, \theta_{z_i}^*)\} \neq \emptyset. \quad \text{(A.7)}$$

Furthermore, $(\{\theta_j^*\}\times D_i\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z)\neq\emptyset$ by the definition of mutually reassuring sets, and it follows from player j's deviation incentives that

$$\mathbf{E}_j^0(u_j(z',\cdot)|(\{\theta_j^*\}\times D_i\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z))>\mathbf{E}_j^0(u_j(z,\cdot)|(\{\theta_j^*\}\times D_i\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z)). \tag{A.8}$$

By the definition of deviating sets,

$$D_i^\beta(\{i,j\},z,z')=\{\theta_i\in\pi_i^{-1}(z):u_i(z',\theta_i,\theta_j^*)>u_i(z,\theta_i,\theta_{z_i}^*)\}. \tag{A.9}$$

Comparing (A.7) and (A.9), we have $D_i^\beta(\{i,j\},z,z')=D_i\neq\emptyset$. The weak consistency of (π,β) requires that

$$\beta_j(\{i,j\},z,z',\theta_j^*)(\cdot)=\beta_j^0(\cdot|(\{\theta_j^*\}\times D_i\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z)). \tag{A.10}$$

Combining (A.8) and (A.10), we obtain $D_j^\beta(\{i,j\},z,z')=\{\theta_j^*\}\neq\emptyset$. Hence,

$$(D_{\{i,j\}}^\beta(\{i,j\},z,z')\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z)=(\{\theta_j^*\}\times D_i\times\Theta_{N\setminus\{i,j\}})\cap\pi^{-1}(z)\neq\emptyset.$$

Therefore, $(\{i,j\},z,z')$ is a viable deviation of (π,β), a contradiction. □

A.5 Duality

This section introduces two results that are consequences of the duality of surplus maximization. Under complete information, the dual of surplus maximization defines pairwise stability. However, under incomplete information, the dual only defines an auxiliary problem where pairwise deviations are not conditional on private information.

Lemma 3. *A playout π is Bayesian efficient if for all $z\in\pi(\Theta)$, $i\in I$, and $j\in J$,*

$$\mathbf{E}^0(u_i(z,\cdot)|\pi^{-1}(z))+\mathbf{E}^0(u_j(z,\cdot)|\pi^{-1}(z)) \geq \mathbf{E}^0(v_i(j,\cdot)+v_j(i,\cdot)|\pi^{-1}(z)); \tag{A.11}$$

$$\mathbf{E}^0(u_i(z,\cdot)|\pi^{-1}(z)) \geq \mathbf{E}^0(v_i(i,\cdot)|\pi^{-1}(z)); \tag{A.12}$$

$$\mathbf{E}^0(u_j(z,\cdot)|\pi^{-1}(z)) \geq \mathbf{E}^0(v_j(j,\cdot)|\pi^{-1}(z)). \tag{A.13}$$

Proof. For any $z\in\pi(\Theta)$, the surplus maximization problem (6.1) in its relaxed form in fractional matching has a dual minimization problem:

$$\min_{(w_n)_{n\in N}}\sum_{n\in N}w_n$$

such that, for any $i\in I$ and $j\in J$,

$$w_i+w_j \geq \mathbf{E}^0(v_i(j,\cdot)+v_j(i,\cdot)|\pi^{-1}(z));$$

$$w_i \geq \mathbf{E}^0(v_i(i,\cdot)|\pi^{-1}(z));$$

$$w_j \;\geq\; \mathbf{E}^0(v_j(j,\cdot)|\pi^{-1}(z)).$$

If conditions (A.11)–(A.13) in Lemma 3 are satisfied, then $(\mathbf{E}^0(u_n(z,\cdot)|\pi^{-1}(z)))_{n\in N}$ is feasible for the dual and hence $\sum_{n\in N}\mathbf{E}^0(u_n(z,\cdot)|\pi^{-1}(z))$ is no less than the optimal value of the primal (6.1). Therefore, z is an optimal solution to the primal. □

Lemma 4. *If the playout π of a weakly consistent and stable assessment (π,β) is not Bayesian efficient, then there exist $z=(\mu,\tau)\in\pi(\Theta)$ and $(i,j,p)\in I\times J\times\mathbb{R}$ such that*

$$\begin{aligned}&\mathbf{E}^0(v_i(j,\cdot)-v_i(\mu_i,\cdot)|\pi^{-1}(z))>\tau_i-p;\\ &\mathbf{E}^0(v_j(i,\cdot)-v_j(\mu_j,\cdot)|\pi^{-1}(z))>\tau_j+p.\end{aligned}\tag{A.14}$$

We call (i,j,p) that satisfies (A.14) an **auxiliary deviation**. It is not a deviation under incomplete information because payoff computations are conditional on neither players' private information nor each other's incentive to deviate.

Proof. Since (π,β) is weakly consistent and stable, the non-existence of any viable scenario of deviation that involves a single player n when the outcome is z implies that

$$\mathbf{E}^0(u_n(z,\cdot)|(\{\theta_n\}\times\Theta_{-n})\cap\pi^{-1}(z))\geq\mathbf{E}^0(v_n(n,\cdot)|(\{\theta_n\}\times\Theta_{-n})\cap\pi^{-1}(z)).$$

Taking an expectation w.r.t. θ_n, we obtain (A.12) and (A.13). If π is not Bayesian efficient, it follows from Lemma 3 that (A.11) is violated for some pair $(i,j)\in I\times J$. Thus, there exists $p\in\mathbb{R}$ such that (A.14) holds. □

A.6 Proof of Theorem 6

Proof. Suppose to the contrary that π is not efficient. By Lemma 4, there exist $p\in\mathbb{R}$, $\bar\theta_i\in\pi_i^{-1}(z)$, and $\bar\theta_j\in\pi_j^{-1}(z)$ such that

$$\mathbf{E}^0(v_i(j,\cdot)|(\{\bar\theta_i\}\times\Theta_{-i})\cap\pi^{-1}(z))+p>\mathbf{E}^0(v_i(\mu_i,\cdot)|(\{\bar\theta_i\}\times\Theta_{-i})\cap\pi^{-1}(z))+\tau_i,\tag{A.15}$$

$$\mathbf{E}^0(v_j(i,\cdot)|(\{\bar\theta_j\}\times\Theta_{-j})\cap\pi^{-1}(z))-p>\mathbf{E}^0(v_j(\mu_j,\cdot)|(\{\bar\theta_j\}\times\Theta_{-j})\cap\pi^{-1}(z))+\tau_j.\tag{A.16}$$

Assume that there is no restriction on v_j by one-sided interdependence. Then (A.15) takes the following form:

$$\mathbf{E}^0(A_i(\cdot)+A_i'(j)|(\{\bar\theta_i\}\times\Theta_{-i})\cap\pi^{-1}(z))+p>\mathbf{E}^0(A_i(\cdot)+A_i'(\mu_i)|(\{\bar\theta_i\}\times\Theta_{-i})\cap\pi^{-1}(z))+\tau_i.$$

Hence, $A_i'(j)+p>A_i'(\mu_i)+\tau_i$. By the definition of deviating sets,

$$D_i^\beta(\{i,j\},z,z')\;=\;\left\{\theta_i\in\pi_i^{-1}(z):\mathbf{E}_{\beta_i(S,z,z',\theta_i)}(u_i(z',\cdot))>\mathbf{E}_{\beta_i(S,z,z',\theta_i)}(u_i(z,\cdot))\right\}$$

$$
\begin{aligned}
&= \left\{\theta_i \in \pi_i^{-1}(z) : \mathbf{E}_{\beta_i(S,z,z',\theta_i)}(v_i(j,\cdot)) + p > \mathbf{E}_{\beta_i(S,z,z',\theta_i)}(v_i(\mu_i,\cdot)) + \tau_i\right\} \\
&= \left\{\theta_i \in \pi_i^{-1}(z) : \mathbf{E}_{\beta_i(S,z,z',\theta_i)}(A_i) + A_i'(j) + p > \mathbf{E}_{\beta_i(S,z,z',\theta_i)}(A_i) + A_i'(\mu_i) + \tau_i\right\} \\
&= \left\{\theta_i \in \pi_i^{-1}(z) : A_i'(j) + p > A_i'(\mu_i) + \tau_i\right\} \\
&= \pi_i^{-1}(z).
\end{aligned}
$$

It follows from the definition of weak consistency that, for any $(\{i,j\}, z, z', \theta_j) \in \Sigma_{\pi,j}$,

$$
\begin{aligned}
\beta_j(\{i,j\}, z, z', \theta_j)(\cdot) &= \beta^0(\cdot|(\{\theta_j\} \times D_i^\beta(\{i,j\}, z, z') \times \Theta_{N\setminus\{i,j\}}) \cap \pi^{-1}(z)) \quad \text{(A.17)} \\
&= \beta^0(\cdot|(\{\theta_j\} \times \Theta_{-j}) \cap \pi^{-1}(z)). \quad \text{(A.18)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
D_j^\beta(\{i,j\}, z, z') &= \left\{\theta_i \in \pi_i^{-1}(z) : \mathbf{E}_{\beta_j(S,z,z',\theta_j)}(u_j(z',\cdot)) > \mathbf{E}_{\beta_j(S,z,z',\theta_j)}(u_j(z,\cdot))\right\} \\
&= \left\{\theta_i \in \pi_i^{-1}(z) : \mathbf{E}_{\beta_j(S,z,z',\theta_j)}(v_j(i,\cdot)) - p > \mathbf{E}_{\beta_j(S,z,z',\theta_j)}(v_j(\mu_j,\cdot)) + \tau_i\right\} \\
&= \left\{\theta_i \in \pi_i^{-1}(z) : \begin{array}{l} \mathbf{E}^0(v_j(i,\cdot)|(\{\theta_j\} \times \Theta_{-j}) \cap \pi^{-1}(z)) - p \\ > \mathbf{E}^0(v_j(\mu_j,\cdot)|(\{\theta_j\} \times \Theta_{-j}) \cap \pi^{-1}(z)) + \tau_i \end{array}\right\} \\
&\neq \emptyset,
\end{aligned}
$$

where the last equality is due to (A.18) and the inequality is due to (A.16). Hence,

$$
(D_{\{i,j\}}^\beta(\{i,j\}, z, z') \times \Theta_{N\setminus\{i,j\}}) \cap \pi^{-1}(z) = (D_j^\beta(\{i,j\}, z, z') \times \Theta_{-j}) \cap \pi^{-1}(z) \neq \emptyset,
$$

contradicting the stability of (π, β). □

A.7 Comonotonicity and Tarski's Fixed Point Theorem

We first prove a mathematical result that follows from comonotonicity.

Lemma 5. *Suppose that $f, g : X_1 \times X_2 \to \mathbb{R}$ are comonotonic[16] on X_1 and X_2, and for some constants c_1 and c_2,*

$$
\mathbf{E}_{\nu_1 \otimes \nu_2}(f) > c_1 \text{ and } \mathbf{E}_{\nu_1 \otimes \nu_2}(g) > c_2, \quad \text{(A.19)}
$$

where the expectation is with respect to some product measure $\nu_1 \otimes \nu_2 \in \Delta(X_1) \times \Delta(X_2)$. Then there exist non-empty sets $D_1^* \subset X_1$ and $D_2^* \subset X_2$ such that

$$
\begin{aligned}
D_1^* &= \{x_1 : \mathbf{E}_{\nu_2}(f|x_1, D_2^*) > c_1\}; \\
D_2^* &= \{x_2 : \mathbf{E}_{\nu_1}(g|x_2, D_1^*) > c_2\}.
\end{aligned} \quad \text{(A.20)}
$$

[16] Here X_3 in the Definition 14 is taken as a singleton set.

The proof idea is as follows. By the comonotonicity of f and g, the mapping defined on the right-hand side of (A.20) is order-reversing in the set-inclusion order, and hence a twice iteration of the mapping is order-preserving and has a fixed point by Tarski's fixed point theorem. A fixed point of the original mapping can be constructed from this fixed point. We then use condition (A.19) to show that the fixed point consists of non-empty sets.

Proof. Suppose without loss of generality that both f and g are non-decreasing with respect to some complete orders $\geq_n$ on X_n. Then consider the class of upper contour sets $B_n(x_n) = \{x'_n : x'_n \geq_n x_n\}$. Let $\mathbb{B}_n = \{B_n(x_n) : x_n \in X_n\} \cup \{\emptyset\}$. Define $d_1 : \mathbb{B}_2 \to 2^{X_1}$ and $d_2 : \mathbb{B}_1 \to 2^{X_2}$ as follows:

$$\begin{aligned} d_1(D_2) &:= \{x_1 : \mathbf{E}_{\nu_2}(f|x_1, D_2) > c_1\}, \;\; d_1(\emptyset) := X_1; \\ d_2(D_1) &:= \{x_2 : \mathbf{E}_{\nu_1}(g|x_2, D_1) > c_2\}, \;\; d_2(\emptyset) := X_2. \end{aligned} \tag{A.21}$$

It follows from $\mathbf{E}_{\nu_1\otimes\nu_2}(f) > c_1$ and $\mathbf{E}_{\nu_1\otimes\nu_2}(g) > c_2$ that $d_1(X_2) \neq \emptyset \neq d_2(X_1)$. Define d on $\mathbb{B}_1 \times \mathbb{B}_2$ as $d(D_1, D_2) = (d_2(D_1), d_1(D_2))$. By monotonicity of f and g, we have $d_1(D_2) \in \mathbb{B}_1$ and $d_2(D_1) \in \mathbb{B}_2$. Therefore d is a self-map on $\mathbb{B}_1 \times \mathbb{B}_2$.

For any $x'_1 \geq_1 x_1$ and $x'_2 \geq_2 x_2$, we have

$$B_1(x'_2) \subset B_1(x_2) \text{ and } B_2(x'_1) \subset B_2(x_1). \tag{A.22}$$

By monotonicity of f and g, we have

$$d_1(B_2(x_2)) \subset d_1(B_2(x'_2)) \text{ and } d_2(B_1(x_1)) \subset d_2(B_1(x'_1)). \tag{A.23}$$

Notice that $\mathbb{B}_1 \times \mathbb{B}_2$ is a complete lattice in the set-inclusion order. It follows from (A.21)–(A.23) that d is order-reversing. Therefore $d^2 : \mathbb{B}_1 \times \mathbb{B}_2 \to \mathbb{B}_1 \times \mathbb{B}_2$ is order-preserving. By Tarski's fixed point theorem, d^2 admits a fixed point (D_1, D_2). By the definitions of d^2 and the fixed point of d^2, we have

$$d^2(D_1, D_2) = d(d_1(D_2), d_2(D_1)) = (d_1(d_2(D_1)), d_2(d_1(D_2))) = (D_1, D_2).$$

Thus $d_1(d_2(D_1)) = D_1$ and hence $(D_1, d_2(D_1))$ is a fixed point of d. The fixed point cannot be of the form $(\emptyset, D)$ because $D = d_2(\emptyset) = X_2$ but $d_1(X_2) \neq \emptyset$. Similarly, the fixed point cannot be of the form $(D, \emptyset)$ because $D = d_1(\emptyset) = X_1$ but $d_Y(X_1) \neq \emptyset$. Therefore, the fixed point of d is non-empty. □

A.8 Proof of Theorem 7

Proof. Suppose to the contrary that π is not Bayesian efficient. By Lemma 4, (A.14) holds for some $z \in \pi(\Theta)$ and a pair of players with transfer (i, j, p). In Lemma 5, set

$$\begin{aligned} X_1 &= \pi_i^{-1}(z),\ X_2 = \pi_j^{-1}(z), \\ f(\theta_i, \theta_j) &= \mathbf{E}^0(v_i(j,\cdot) - v_i(\mu_i,\cdot) | (\{\theta_i\} \times \{\theta_j\} \times \Theta_{N\setminus\{i,j\}}) \cap \pi^{-1}(z)), \\ g(\theta_i, \theta_j) &= \mathbf{E}^0(v_j(i,\cdot) - v_j(\mu_j,\cdot) | (\{\theta_i\} \times \{\theta_j\} \times \Theta_{N\setminus\{i,j\}}) \cap \pi^{-1}(z)), \\ c_1 &= \tau_i - p,\ c_2 = \tau_j + p, \\ \nu_1(\theta_i) &= \beta^0(\theta_i | \pi^{-1}(z)),\ \nu_2(\theta_j) = \beta^0(\theta_j | \pi^{-1}(z)). \end{aligned}$$

Since the playout π has independent types, f and g are well-defined on $\pi_i^{-1}(z) \times \pi_j^{-1}(z)$. Since the game has comonotonic differences, f and g are comonotonic on $\pi_i^{-1}(z)$ and $\pi_j^{-1}(z)$. Condition (A.14) implies (A.19), and hence there exist non-empty sets D_1^* and D_2^* that satisfy (A.20). They are mutually reassuring sets for some deviation $(\{i,j\}, z, z')$ in which i and j match with transfer p in z'. By Theorem 2, the existence of mutually reassuring sets contradicts with the assumption that (π, β) is strongly consistent and stable. □

A.9 Proof of Theorem 8

We first flesh out the extension in our network setting. A deviation in a network takes the form $\delta = (\{m,n\}, z, z', s_n, s_m)$. A perceived scenario of deviation of player n takes the form of $\sigma_n = (\{m,n\}, z, z', s_n, s_m, \theta_{z_n})$, because player n observes not only his own type θ_n, but also the types of all players in z_n. Let $\Sigma_{\pi,n}$ be the set of player n's perceived scenarios of deviation. A *belief system* for a playout π is $\beta = (\beta_n)_{n\in N}$, where $\beta_n : \Sigma_{\pi,n} \to \Delta(\Theta)$ is such that for any $\sigma_n \in \Sigma_{\pi,n}$ we have $\beta_n(\sigma_n)(\{\theta_{z_n}\} \times \Theta_{-n}) = 1$ and $\beta_n(\sigma_{z_n})(\pi^{-1}(z)) = 1$. Consider a deviation δ, the *deviation set* for player n is

$$D_n^\beta(\delta) := \left\{\theta_{z_n} \in \pi_{z_n}^{-1}(z) : \mathbf{E}_{\beta_n(\delta,\theta_{z_n})}(u_n(z',\cdot) - C_{nm}(s_n,\cdot)) > \mathbf{E}_{\beta_n(\delta,\theta_{z_n})}(u_n(z,\cdot))\right\}.$$

Weak consistency says that $\beta_n(\delta, \theta_{z_n})(\cdot) = \beta_n^0(\cdot | (\{\theta_{z_n}\} \times \Theta_{-z_n}) \cap (D_m^\beta(\delta) \times \Theta_{-z_m}) \cap \pi^{-1}(z))$.

We say $D_n \subset \pi_{z_n}^{-1}(z)$ and $D_m \subset \pi_{z_m}^{-1}(z)$ are *mutually reassuring with certainty* for a deviation $\delta = (\{m,n\}, z, z', s_m, s_n)$, where $m \notin z_n$, if the following two conditions hold:

(i) $(D_m \times \Theta_{-z_m}) \cap (D_n \times \Theta_{-z_n}) \cap \pi^{-1}(z) \neq \emptyset$;

(ii) D_m and D_n satisfy the following fixed-point property:

$$D_m = \left\{ \theta_{z_m} \in \pi_{z_m}^{-1}(z) : \begin{array}{c} \text{(a) } (\{\theta_{z_m}\} \times \Theta_{-z_m}) \cap (D_n \times \Theta_{-z_n}) \cap \pi^{-1}(z)) \neq \emptyset \\ \text{(b) } \mathbf{E}_m^0(u_m(z', \cdot) - C_{mn}(s_m, \cdot) | (\{\theta_{z_m}\} \times \Theta_{-z_m}) \cap (D_n \times \Theta_{-z_n}) \cap \pi^{-1}(z)) \\ > \mathbf{E}_m^0(u_m(z, \cdot) | (\{\theta_{z_m}\} \times \Theta_{-z_m}) \cap (D_n \times \Theta_{-z_n}) \cap \pi^{-1}(z)) \end{array} \right\},$$

$$D_n = \left\{ \theta_{z_n} \in \pi_{z_n}^{-1}(z) : \begin{array}{c} \text{(a) } (\{\theta_{z_n}\} \times \Theta_{-z_n}) \cap (D_m \times \Theta_{-z_m}) \cap \pi^{-1}(z)) \neq \emptyset \\ \text{(b) } \mathbf{E}_n^0(u_n(z', \cdot) - C_{nm}(s_n, \cdot) | (\{\theta_{z_n}\} \times \Theta_{-z_n}) \cap (D_m \times \Theta_{-z_m}) \cap \pi^{-1}(z)) \\ > \mathbf{E}_n^0(u_n(z, \cdot) | (\{\theta_{z_n}\} \times \Theta_{-z_n}) \cap (D_m \times \Theta_{-z_m}) \cap \pi^{-1}(z)) \end{array} \right\}.$$

An assessment (π, β) is *strongly consistent with certainty* if (i) it is weakly consistent and (ii) a deviation $\delta = (\{m, n\}, z, z', s_m, s_n)$, where $m \notin z_n$, is viable if there exist D_m and D_n that are mutually reassuring with certainty for the deviation.

The definitions of mutual reassurance, conscious conditional probability, and strong consistency can be extended analogously.

Proof of Theorem 8. Consider an assessment (π, β). Suppose to the contrary that there exists θ^* such that $z = \pi(\theta^*)$ is *not* a stable outcome for Γ^{θ^*}. Then there are two cases to consider. First, there exist $n \neq m \in z_n$ such that under complete information of θ^*, player n will remove player m from z_n. That is,

$$v_{nm}(\theta_n^*, \theta_m^*) < c_n(|z_n|, \theta_n^*) - c_n(|z_n| - 1, \theta_n^*).$$

Since player n observes player m's type θ_m^* under incomplete information, he will remove m from z_n as well.

Second, there exists $n \neq m \notin z_n$ such that both n and m would like to connect with each other under complete information of θ^*. We define

$$\Delta_{mn}(\theta_m, \theta_n) := v_{mn}(\theta_m, \theta_n) - (c_m(|z_m| + 1, \theta_m) - c_m(|z_m|, \theta_m))$$

as the net benefit of forming a link with n for player m. It follows from Assumption 1 that Δ_{mn} is strictly increasing. We define $\Delta_{nm}(\theta_n, \theta_m)$ similarly. By the choice of (m, n) and θ^*, we have $\Delta_{mn}(\theta_m^*, \theta_n^*) > 0$ and $\Delta_{nm}(\theta_n^*, \theta_m^*) > 0$.

Choose any $\theta^{**} \in \pi^{-1}(z)$ such that $(\Delta_{mn}(\theta_m^{**}, \theta_n^{**}), \Delta_{nm}(\theta_n^{**}, \theta_m^{**}))$ is Pareto undominated over all θ's in $\pi^{-1}(z)$. Choose $\epsilon > 0$ such that $\Delta_{mn}(\theta_m^{**}, \theta_n^{**}) - \epsilon > 0$ and $\Delta_{nm}(\theta_n^{**}, \theta_m^{**}) - \epsilon > 0$, and meanwhile, for any $\theta_m < \theta_m^{**}$ and $\theta_n < \theta_n^{**}$, $\Delta_{mn}(\theta_m^{**}, \theta_n^{**}) - \Delta_{mn}(\theta_m, \theta_n^{**}) > \epsilon$ and $\Delta_{nm}(\theta_n^{**}, \theta_m^{**}) - \Delta_{nm}(\theta_n, \theta_m^{**}) > \epsilon$. The existence of ϵ is guaranteed because Δ_{mn} and Δ_{nm} are strictly increasing.

By Assumption 1, there exist $s_m, s_n \in \mathbb{R}_+$ such that $C_{mn}(s_m, \theta_m^{**}, \theta_n^{**}) = \Delta_{mn}(\theta_m^{**}, \theta_n^{**}) - \epsilon$ and $C_{nm}(s_n, , \theta_n^{**}, \theta_m^{**}) = \Delta_{nm}(\theta_n^{**}, \theta_m^{**}) - \epsilon$. We claim that there exists a unique θ_m, namely θ_m^{**}, such that the following holds:

$$\Delta_{mn}(\theta_m, \theta_n^{**}) - C_{mn}(s_m, \theta_m, \theta_n^{**}) > 0, \tag{A.24}$$

$$(\{\theta_m\} \times \Theta_{-m}) \cap (\{\theta_n^{**}\} \times \Theta_{-n}) \cap \pi^{-1}(z) \neq \emptyset. \tag{A.25}$$

To see this, note that θ_m^{**} satisfies (A.24) and (A.25) by the choice of s_m and θ_m^{**}. If $\theta_m > \theta_m^{**}$ satisfies both (A.24) and (A.25), then there is a contradiction with the Pareto undominance of $(\Delta_{mn}(\theta_m^{**}, \theta_n^{**}), \Delta_{nm}(\theta_n^{**}, \theta_m^{**}))$. If $\theta_m < \theta_m^{**}$, then

$$\begin{aligned}
\Delta_{mn}(\theta_m, \theta_n^{**}) - C_{mn}(s_m, \theta_m, \theta_n^{**}) &\leq \Delta_{mn}(\theta_m, \theta_n^{**}) - C_{mn}(s_m, \theta_m^{**}, \theta_n^{**}) && \text{(A.26)} \\
&= \Delta_{mn}(\theta_m, \theta_n^{**}) - \Delta_{mn}(\theta_m^{**}, \theta_n^{**}) + \epsilon && \text{(A.27)} \\
&< 0, && \text{(A.28)}
\end{aligned}$$

where (A.26) is due to the monotonicity of C_{mn} in θ_m, (A.27) is by the choice of s_m, and (A.28) is by the choice of ϵ. This contradicts (A.24).

Similarly, θ_n^{**} is the only θ_n that satisfies $\Delta_{nm}(\theta_n, \theta_m^{**}) - C_n(s_n, \theta_n, \theta_m^{**}) > 0$ and $(\{\theta_n\} \times \Theta_{-n}) \cap (\{\theta_m^{**}\} \times \Theta_{-m}) \cap \pi^{-1}(z) \neq \emptyset$.

Let us define $D_m := \{\theta_{z_m} \in \pi_{z_m}^{-1}(z) : \theta_m = \theta_m^{**}\}$ and $D_n := \{\theta_{z_n} \in \pi_{z_n}^{-1}(z) : \theta_n = \theta_n^{**}\}$. Therefore,

$$\begin{aligned}
&\mathbf{E}_m^0(u_m(z', \cdot) - u_m(z, \cdot) - C_{mn}(s_m, \cdot) | (\{\theta_{z_m}\} \times \Theta_{-z_m}) \cap (D_n \times \Theta_{-z_n}) \cap \pi^{-1}(z)) \\
&= \Delta_{mn}(\theta_m, \theta_n^{**}) - C_{mn}(s_m, \theta_m, \theta_n^{**}).
\end{aligned}$$

An analogous expression holds for n. Because θ_m^{**} is the only θ_m that satisfies (A.24) and (A.25), D_m and D_n are mutually reassuring with certainty. □

References

Aumann, Robert J and Aviad Heifetz (2002). "Incomplete information". In: *Handbook of Game Theory with Economic Applications* 3, pp. 1665–1686.

Bikhchandani, Sushil (2017). "Stability with one-sided incomplete information". In: *Journal of Economic Theory* 168, pp. 372–399.

Chen, Yi-Chun and Gaoji Hu (2023). "A theory of stability in matching with incomplete information". In: *American Economic Journal: Microeconomics* 15.1, pp. 288–322.

Crawford, Vincent P and Elsie Marie Knoer (1981). "Job matching with heterogeneous firms and workers". In: *Econometrica*, pp. 437–450.

Dizdar, Deniz and Benny Moldovanu (2016). "On the importance of uniform sharing rules for efficient matching". In: *Journal of Economic Theory* 165, pp. 106–123.

Dubey, Pradeep and Lloyd S Shapley (1984). "Totally balanced games arising from controlled programming problems". In: *Mathematical Programming* 29, pp. 245–267.

Dutta, Bhaskar and Suresh Mutuswami (1997). "Stable networks". In: *Journal of Economic Theory* 76.2, pp. 322–344.

Dutta, Bhaskar and Rajiv Vohra (2005). "Incomplete information, credibility and the core". In: *Mathematical Social Sciences* 50.2, pp. 148–165.

Fisher, Franklin M (1991). "Organizing industrial organization: reflections on the handbook of industrial organization". In: *Brookings Papers on Economic Activity. Microeconomics* 1991, pp. 201–240.

Forges, Françoise (1994). "Posterior efficiency". In: *Games and Economic Behavior* 6.2, pp. 238–261.

Forges, Françoise, Enrico Minelli, and Rajiv Vohra (2002). "Incentives and the core of an exchange economy: a survey". In: *Journal of Mathematical Economics* 38.1-2, pp. 1–41.

Gale, David and Lloyd S Shapley (1962). "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1, pp. 9–15.

Green, Jerry R and Jean-Jacques Laffont (1987). "Posterior implementability in a two-person decision problem". In: *Econometrica*, pp. 69–94.

Gul, Faruk (1989). "Bargaining foundations of Shapley value". In: *Econometrica*, pp. 81–95.

Gul, Faruk and Wolfgang Pesendorfer (2020). "Lindahl equilibrium as a collective choice rule". In: *Working Paper, Princeton University*.

Holmström, Bengt and Roger B Myerson (1983). "Efficient and durable decision rules with incomplete information". In: *Econometrica*, pp. 1799–1819.

Jackson, Matthew O and Asher Wolinsky (1996). *A strategic model of social and economic networks*. Vol. 71. 1, pp. 44–74.

Kamishiro, Yusuke, Rajiv Vohra, and Roberto Serrano (2022). "Signaling, Screening, and Core Stability". In: *Working Paper, Brown University*.

Kreps, David M and Robert Wilson (1982). "Sequential equilibria". In: *Econometrica*, pp. 863–894.

Levin, Jonathan (2003). "Lecture Notes on Game Theory". In: *Stanford University*.

Liu, Qingmin (2020). "Stability and Bayesian consistency in two-sided markets". In: *American Economic Review* 110.8, pp. 2625–2666.

Liu, Qingmin, George J Mailath, Andrew Postlewaite, and Larry Samuelson (2014). "Stable matching with incomplete information". In: *Econometrica* 82.2, pp. 541–587.

Milgrom, Paul and Nancy Stokey (1982). "Information, trade and common knowledge". In: *Journal of Economic Theory* 26.1, pp. 17–27.

Okada, Akira (2012). "Non-cooperative bargaining and the incomplete informational core". In: *Journal of Economic Theory* 147.3, pp. 1165–1190.

Osborne, Martin J and Ariel Rubinstein (1994). *A course in game theory*. MIT press.

Perry, Motty and Philip J Reny (1994). "A noncooperative view of coalition formation and the core". In: *Econometrica*, pp. 795–817.

Ray, Debraj (2007). *A game-theoretic perspective on coalition formation*. Oxford University Press.

Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information". In: *The Quarterly Journal of Economics* 90.4, pp. 629–649.

Rubinstein, Ariel and Asher Wolinsky (1985). "Equilibrium in a market with sequential bargaining". In: *Econometrica: Journal of the Econometric Society*, pp. 1133–1150.

— (1990). "On the logic of "agreeing to disagree" type results". In: *Journal of Economic Theory* 51.1, pp. 184–193.

Sadler, Evan (2022). "Making a Swap: Network Formation with Increasing Marginal Costs". In: *Working Paper, Columbia University*.

Shapley, Lloyd S (1967). "On balanced sets and cores". In: *Naval Research Logistics Quarterly* 14, pp. 453–460.

Shapley, Lloyd S and Martin Shubik (1971). "The assignment game I: The core". In: *International Journal of Game Theory* 1.1, pp. 111–130.

Wilson, Robert (1978). "Information, efficiency, and the core of an economy". In: *Econometrica*, pp. 807–816.