NORTHWESTERN UNIVERSITY

Essays in Microeconomic Theory: Information Design, Partnerships, and Matching

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Béla Szabadi

EVANSTON, ILLINOIS

September 2018

Abstract

This dissertation consists of three essays in Microeconomic Theory that study the interplay between mechanism and information design, provide insights into the design of efficient dispute resolution mechanisms for partnerships, and analyze the stability of fractional matchings in two-sided markets.

In the first chapter, we study the optimal disclosure policy in a setting in which both a monopolistic seller and a single potential buyer of an indivisible good face product-related uncertainty. The buyer's valuation depends on his privately known type and the unknown attributes of the object. The seller designs a disclosure policy with a publicly observable outcome and interacts with the buyer only after the information is revealed. We identify the condition on the underlying preferences that guarantees the optimality of full disclosure: if buyer types are ranked uniformly across states by their willingness to pay, the seller prefers to disclose all information about her product. Whenever this condition is violated, there always exist type distributions for which full disclosure is suboptimal. We also study the optimal information disclosure policy in cases in which the released information might affect the ranking of the types. In environments with two types, the seller prefers partial disclosure that never reverses the ranking and fully reveals states in which the net gain from restricting output by selling only to the higher type is above a certain threshold. In settings with many types, the optimal disclosure policy depends on the distribution of types in a more complicated way. We establish structural properties that the groups of types that purchase the object at different posteriors must fulfill under the optimal policy.

The second chapter, which is a joint work with Daniel Fershtman, studies the efficient resolution of partnership disputes where efficiency need not imply dissolution. The modeled

endogeneity of dissolution is a departure from the partnership dissolution literature emanating from Cramton et al. (1987), which has effectively focused on circumstances in which dissolution is unavoidable; the analysis yields predictions that contrast sharply with their analogs in such environments. First, we show that unless a dispute is sufficiently severe, its efficient resolution is infeasible without external subsidy. Furthermore, we prove that the severity of a dispute has a non-monotonic effect on the cost of its efficient resolution. We also consider an alternative class of mechanisms and characterize the disputes for which a profit-oriented arbitrator benefits from a more conservative approach in opting for dissolution. The latter characterization has implications for the design of arbitration in environments in which the partners' decision to trigger a dispute is itself endogenous.

The third chapter describes a new stability concept for fractional matchings in two-sided markets in which agents can be matched with multiple partners over time. The new definition, called C-stability, is equivalent to the non-existence of unstable implementing time schedules and is less restrictive than strong stability proposed by Roth et al. (1993). We characterize this condition both geometrically and in terms of the underlying preferences. The geometric characterization reveals the relationship between the set of fractional matchings satisfying C-stability, the polytope of matchings, and the polytope of weakly stable fractional matchings. The preference-based characterization uses these results to provide necessary and sufficient conditions on the underlying preferences that a fractional matching must fulfill to satisfy C-stability. These findings also highlight the difference between strong stability and C-stability and identify the exact cases in which strong stability excludes matchings that can be implemented only by stable time schedules.

Acknowledgments

I would like to express my deepest gratitute to my advisors, Asher Wolinsky, Wojciech Olszewski, and Alessandro Pavan for their guidance, encouragement and support during my PhD studies at Northwestern University.

I am also grateful to Daniel Fershtman for being a fantastic coauthor. The different chapters also benefited a lot from discussions with Arjada Bardhi, Eddie Dekel, Jeffrey Ely, Nicolas Inostroza, James Schummer, Marciano Siniscalchi, Bruno Strulovici, and Rainer Widmann. I also thank Judith Levi for her excellent editorial work on the paper that served as the basis for the first chapter. I am grateful to the Center for Economic Theory for the financial support.

I also thank my former professors whose devotion to science and continuous encouragement led me to pursue a doctoral degree in economics: Larry Blume, Egbert Dierker, Christian Haefke, Tamás Papp, and Klaus Ritzberger from the Institute for Advanced Studies Vienna, and Imre Csekő, Gyula Magyarkuti, and Ernő Zalai from the Corvinus University of Budapest.

Last but not least, I am deeply thankful to my wife Anikó, my parents Béla and Ilona, and my close friends. They were always there for me to provide the emotional support I needed during my journey through graduate school.

Contents

List of Figures

List of Tables

Chapter 1

Optimal Disclosure with Subsequent Screening

1.1 Introduction

There are many economic situations in which one party has the option to reveal information that, in a later strategic interaction, might be relevant to all participating parties, including herself. An online seller can design a customer-review system that not only provides information to potential buyers about the product, but also helps her assess the product's reception by the consumers. A university's teaching evaluation system informs the institution, its instructors, and its students about the quality of the education in a way that might have complex effects on the relationship between these groups; for example, later promotion and hiring decisions can rely on the data, which might also shape or misshape the priorities of the instructors in teaching. In an acquisition process, before negotiating the terms, the executives of a target tech company may have to decide how detailed an investigation they will allow into the company's intellectual property and financial situation. In all of these cases, the information disclosure policy has to be designed with careful consideration of its later effects on the strategic interaction.

This chapter studies such an optimal disclosure problem in a setting in which both a monopolistic seller and a potential buyer face uncertainty about a product. The buyer's valuation is influenced by both his privately known type and the unknown product characteristics. The seller designs an experiment that reveals a publicly observed signal about these unknown characteristics before interacting with the buyer. The provided information might have heterogeneous effects on different types that are strong enough to change their ranking according to their willingness to pay. Our goal is to describe the seller's optimal disclosure policy and its dependence on the underlying model parameters in this environment.

We show that the key feature of the seller's design problem is the relationship between the released information and its effect on the ranking of the types according to their willingness to pay. If the released information cannot change the ranking, then it is optimal for the seller to use a fully revealing disclosure policy to be able to condition the group of types to which she sells and the price she charges on the different states. Whenever this condition is violated, there always exist type distributions for which full disclosure is suboptimal. In such cases, the seller can capitalize on the heterogeneous effects of the revealed information by leaving uncertainty between some states that induce different rankings to equalize the expected valuations of some types and capture more expected surplus.

This work also takes a step further by analyzing the optimal disclosure policy in cases in which the released information can potentially change the ranking of the types. We provide a complete characterization for environments with two buyer types. In such settings, no matter what the underlying type distribution is, fully disclosing the actual state is never optimal for the seller if types are not ranked in the same way across states. The optimal policy discloses information that either (i) fully reveals states in which the order of the types is consistent with the prior order, or (ii) induces a belief that assigns positive probability to two states that correspond to different rankings in a way that equalizes the expected valuations. In particular, the seller prefers to reveal states in which her net gain from restricting the output by selling only to the higher type instead of selling to both types is above a certain threshold.

In the two-type environment, we also decompose the original state space into a two-dimensional structure that separates the horizontal and vertical effects of the released information. The horizontal dimension captures the part of the information that can affect the ranking while the vertical information never changes the order conditional on the horizontal state. Under the optimal information disclosure policy, information about the vertical dimension is fully revealed, while information along the horizontal dimension is only partially disclosed.

Still in the two-type environment, we also relate the welfare consequences of the optimal disclosure policy to those in the benchmark cases of no disclosure and full disclosure. When it is compared to no disclosure, the effects of the optimal information disclosure on the probability of sale, and consequently on the social welfare and consumer surplus, are ambiguous. Compared to full disclosure, however, the optimal policy induces a more efficient outcome, but more than this efficiency gain is captured by the seller, leading to a lower consumer surplus.

In settings with many buyer types, the optimal disclosure policy depends on the underlying type distribution in a more complicated way. We provide a partial characterization of the optimal policy that generalizes multiple results from the two-type environment. First, the released information either fully reveals the true state or induces a belief that equalizes the expected valuations of *at least* two types. Second, optimality dictates that certain properties must be satisfied by the groups of types that may purchase the product depending on the revealed information. Every pair of beliefs that arise with positive probability under the optimal policy has the following three properties. First, the groups of types that purchase the good given each of these beliefs cannot be disjoint. Second, if the first group has a mass weakly larger than that of the second one, then the marginal types in the first group cannot all receive the product under the second belief and be non marginal types. And third, if the first group has a strictly larger mass, then the marginal types who purchase given the first belief cannot all receive the good under the second belief. These properties generalize the results derived in the two-type environment, namely, that the released information cannot reverse the ranking of types, and that full disclosure must be suboptimal whenever types are not uniformly ordered across states.

1.1.1 Related literature

Many papers analyze a seller's optimal information disclosure problem in selling or auction environments (see, for example, Milgrom and Weber (1982), Lewis and Sappington (1994), Ottaviani and Prat (2001), Eső and Szentes (2007), Board (2009), Fu et al. (2012)). Although these papers use different modeling assumptions, their general conclusion is that one of the two extreme information disclosure policies—full disclosure or no disclosure—is optimal for the seller. Heterogeneity between different bidders does not significantly alter this conclusion: although releasing all information before a second price auction can be detrimental to the seller's revenue if the information can affect *the ranking of the bidders* (the allocation effect in Board (2009)), the optimality of full disclosure is restored if the auction format can be chosen conditional on the revealed information (Fu et al. (2012)). We contribute to this literature by presenting a model in which a different kind of heterogeneity is allowed, and the revealed information might have strong heterogeneous effects on the *different types* of a given buyer. In a model in which disclosure precedes interaction with the buyer, we show that this heterogeneity can give an incentive for the seller to only partially reveal the underlying uncertainty.

Such *taste heterogeneity* between different buyer types is also present in the papers of Koessler and Renault (2012) and Koessler and Skreta (2016). These works, however, consider environments in which the seller has private information about her own type. Koessler and Renault (2012) focus on a strategic communication setting in which the seller can credibly disclose her private information. In their model, the unraveling of seller types (and hence full disclosure) is always an equilibrium, and in many cases it is the unique one. In contrast, the present work assumes that the underlying state is unknown even to the seller at the information design stage, and she can disclose the information with commitment. Consequently, full revelation will happen less frequently than in their model. Koessler and Skreta (2016) consider an informed principal problem and analyze the selling procedures that can arise in equilibrium.

Li and Shi (2017a) and Li and Shi (2017b) present a different type of single-buyer model in which partial disclosure can be optimal for the seller. In their setting, the seller can interact with the buyer before disclosure, and the released information and its effect on the buyer's valuation depend on both the reported and the underlying true type of the buyer.

In a recent work, for environments in which types and states might be correlated, Yamashita (2018) shows that independence guarantees the optimality of full disclosure (similarly to the uniform ranking condition in the first statements of Propositions 1.1 and 1.5 in the present work). The paper offers counterexamples for the converse that illustrate the potential suboptimality of full disclosure, while the present work provides more general results by showing that suboptimality is guaranteed for a non-trivial set of type distributions.

Condorelli and Szentes (2017) and Roesler and Szentes (2017) consider the information design problem in a buyer-seller setting from the buyer's point of view. In these works, the buyer chooses either the distribution of his valuations (Condorelli and Szentes (2017)) or the distribution of a signal of his underlying true valuation (Roesler and Szentes (2017)). In both models, the seller observes only the distribution chosen but not its realization before making a take-it-or-leave-it price offer to the buyer. Contrary to this, in the present model, a signal about the underlying state is designed by the seller and observed by both parties before the interaction.

Methodologically, the model in this chapter is related to the literature on Bayesian Persuasion, especially to the possible generalizations mentioned in Kamenica and Gentzkow (2011). The first of these generalizations is that the buyer has private information, and the second is that the model can be interpreted as one with multiple receivers. The monopolist sender also learns from and bases later decisions on the information structure she has designed. However, this second-stage strategic interaction between the seller and the buyer can be reduced to a decision problem of the seller in which she chooses whether to charge the lower or the higher expected valuation based on the revealed information. Therefore, in the reduced form of our model, the concavification argument of Kamenica and Gentzkow (2011) is still valid.

Alonso and Câmara (2016) study a persuasion problem in which a politician aims to convince a fraction of voters to support a proposal. In the special case when a single voter has to be persuaded, the proposal is approved with certainty in states in which the voter's net gain from it is above a certain cutoff, and is rejected otherwise. We derive a similar feature of the optimal information structure in the two-type case: the seller optimally restricts output and sells only to the higher type exactly in the states in which the net gain from doing so is above a certain threshold. However, the underlying setting, the linear programming problem establishing this result, and the geometric intuition behind it are different in our model. In particular, the receiver (who is also the seller in the reduced form of the problem) does not necessarily have to be indifferent between the two alternatives (i.e., charging type 1's or type 2's valuation) when any of these alternatives is chosen.

Finally, the present work is also related to management science papers on consumer reviews and information disclosure, such as Jiang and Guo (2015), Zhang et al. (2017), Kwark et al. (2014), and Li et al. (2011). These papers typically distinguish between uncertainty about vertical and horizontal product attributes. Vertical attributes (e.g., build quality) are perceived by all consumer types in the same way; they typically induce parallel shifts in the demand function. In contrast, horizontal characteristics (e.g., misfit costs) have heterogeneous effects on the different consumer groups. These papers are concerned with the effects of review systems and information disclosure in given market structures, depending on the relative importance of horizontal and vertical characteristics in the buyers' utility functions. In particular, Jiang and Guo (2015) analyze the design of review systems in a restricted information-design environment in which consumers, after consuming a good, pick a category rating based on their utility levels. The number of available rating levels is determined by the seller. They show that if vertical attributes are more important (i.e., if misfit costs tend to be small), the seller prefers to provide the finest rating system possible, and the coarsest one otherwise. We expand the theoretical insights of these papers in two ways. First, we use less restrictive definitions for vertical and horizontal attributes; the vertical dimension of information can also have heterogeneous effects on types and can even move their valuations in the opposite direction as long as it does not change their ranking. Second, given this structure, we show that in an unrestricted information-design environment, the seller prefers to disclose the vertical dimension of information completely and typically favors partial disclosure for the horizontal dimension.

The chapter is organized as follows. In Section 1.2, we formally present the information disclosure and selling environment. Section 1.3 describes the results for two possible types. Section 1.3.1 identifies the condition that guarantees the optimality of full disclosure, and Section 1.3.2 describes the optimal disclosure policy for cases when this condition is violated.

Section 1.3.3 presents the horizontal-vertical decomposition of the state space and Section 1.3.4 discusses the welfare consequences of the optimal disclosure policy. Section 1.4 presents the generalized results about the optimal disclosure policy in the many-type environment. Section 1.5 concludes the chapter. Appendix A contains the proofs omitted from the main text.

1.2 Model

Consider an environment in which a monopolistic seller *(she)* wants to sell a single good to a buyer *(him)* with unit demand such that both the buyer and the seller face some product-related uncertainty at the beginning of the game. This uncertainty about the seller's product is described by a finite state space $\Psi = \{\psi_1, \ldots, \psi_K\}$. We assume that the seller and the buyer have the same prior belief about the product, which we denote by the vector $\boldsymbol{\rho} = \{\rho_1, \ldots, \rho_K\} \in \operatorname{int} \Delta^{K-1}$.[1]

The buyer's willingness to pay depends on the unknown characteristics of the product $\psi \in \Psi$ and his privately known taste parameter $\theta \in \Theta \doteq \{\theta_1, \ldots, \theta_I\}$. The seller knows only the distribution of the buyer's taste type, denoted by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I) \in \operatorname{int} \Delta^{I-1}$.[2] Type θ_i's willingness to pay in state ψ_k is denoted by v_{ik}. For each i, the vector $\boldsymbol{v_{i\cdot}} \in \mathbb{R}^K_{++}$ represents valuations of type θ_i across states. Besides assuming that $v_{ik} > 0$ for each i and k, we do not place any restriction on the valuations. Most importantly, the released information may have the potential to change the ranking of the types in an arbitrary way, which reflects the fact that information might have strong heterogeneous effects on different buyer types.

[1]Vectors in the dissertation are written in boldface. The inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of the same dimension is denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.

[2]Although the buyer's type is also part of the uncertainty the seller faces, we still use the term *state* to refer only to the product-related uncertainty.

Marginal cost is assumed to be zero in every state.[3] At the beginning of the game, the seller chooses the information disclosure device and commits to it. We assume that screening and contracting must follow the disclosure of information, and we analyze the design of the optimal information disclosure policy with its effects on the later mechanism design environment in mind.[4]

The information disclosure system is modelled in a standard way: it consists of a set of signal realizations $Z = \{z_1, \ldots, z_L\}$, and for each state $\psi_k \in \Psi$, a distribution over the signal realizations $\boldsymbol{\zeta}^{(k)} \in \Delta^{L-1}$ conditional on the the state being ψ_k. By changing the rules of the information disclosure system, the seller can influence the quantity and the quality of the information that might be revealed about the product.

The timing of the game can be summarized as follows:

1. The seller designs the disclosure policy $\left(Z, (\boldsymbol{\zeta}^{(k)})_{k=1}^{K}\right)$.
2. A signal is generated.
3. The seller and the buyer learn the signal realization and update their beliefs using Bayes' rule.
4. The seller offers a selling mechanism to the buyer.
5. The buyer decides whether or not to participate.

[3]The model can be easily generalized to include a *state-independent*, positive marginal cost c in settings where it is profitable to sell the good for every posterior belief. Most of the results presented here would generalize to this case with the redefined valuations $v'_{ik} = \max\{v_{ik} - c, 0\}$. On the other hand, the presence of a *state-dependent* marginal cost in the model would most likely create more incentive for the seller to disclose information since she could avoid selling the product to types whose valuation is lower than her own marginal cost. We believe that in real-life information design problems, the latter should be a less important factor in the seller's decision than the effect on the valuations of the consumer groups. For this reason and also to ease the exposition, we do not include this additional trade-off in the model.

[4]This also means that the buyer's individual rationality constraint must be satisfied only *after* the information is revealed, and hence it does not matter whether or not the seller commits to the selling mechanism at the beginning of the game.

In the two-type environment, the model allows us to draw very clear conclusions about the optimal information structure and its welfare implications. Therefore, we study this special case first before presenting the general results.

1.3 Results for two types

In this section, we describe the optimal disclosure policy in the two-type environment ($I = 2$). Let $\boldsymbol{\alpha} = (\alpha, 1-\alpha)$, where $\alpha > 0$, be the distribution of the two types. First, we formally define the seller's optimal information design problem.

We know that in environments like the one presented here, the problem of designing signals is equivalent to choosing a distribution of posterior beliefs satisfying the law of total probability.[5] Denote such a distribution by $(\boldsymbol{\rho}^{(1)}, \dots, \boldsymbol{\rho}^{(P)}; \boldsymbol{\lambda}) \in \Delta^{K-1} \times \dots \times \Delta^{K-1} \times \Delta^{P-1}$, where λ_p is the probability that posterior $\boldsymbol{\rho}^{(p)}$ is reached. The law of total probability requires that the expected posterior belief be equal to the prior $\boldsymbol{\rho}$:

$$\sum_{p=1}^{P} \lambda_p \boldsymbol{\rho}^{(\boldsymbol{p})} = \boldsymbol{\rho}.$$

A particularly important posterior distribution corresponds to the fully revealing signal, denoted by $(\boldsymbol{e}^{(\mathbf{1})}, \dots, \boldsymbol{e}^{(\boldsymbol{K})}; \boldsymbol{\rho})$, where $\boldsymbol{e}^{(\boldsymbol{k})}$ is the k-th unit vector in $\mathbb{R}^K$.

Given the setting of a single buyer with unit demand, offering a take-it-or-leave-it price conditional on the disclosed information is optimal for the seller, and so our attention will be restricted to this class of selling mechanisms from now on. To maximize profits, the seller optimally chooses between selling to both types at the lower type's expected valuation or only to the higher type at the higher type's expected valuation. Thus, the maximal (indirect)

[5] See, for example, Kamenica and Gentzkow (2011).

profit she can attain at any posterior belief $\tilde{\boldsymbol{\rho}}$ is given by the following formula,

$$\Pi(\tilde{\boldsymbol{\rho}};\alpha) \doteq \max\left\{\min\left\{\langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{1\cdot}}\rangle,\langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{2\cdot}}\rangle\right\},\alpha\langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{1\cdot}}\rangle,(1-\alpha)\langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{2\cdot}}\rangle\right\},$$

where one of the last two terms inside the maximum function represents the profit from the latter "discriminatory" selling policy (the one for which the expected valuation is higher), while the other term is dominated by the minimum of the two expected valuations.

The seller's goal is to find posteriors $(\boldsymbol{\rho}^{(1)},\ldots,\boldsymbol{\rho}^{(P)};\boldsymbol{\lambda})$ that satisfy the law of total probability and maximize the expected profit given the indirect profit function defined above. Formally,

$$\begin{aligned}\max_{\substack{P\in\mathbb{N},\\ \boldsymbol{\rho^{(p)}}\in\Delta^{K-1}\ \forall p,\\ \boldsymbol{\lambda}\in\Delta^{P-1}}} \quad & \sum_{p=1}^{P}\lambda_p\Pi(\boldsymbol{\rho^{(p)}};\alpha) \\ \text{s.t.} \quad & \sum_{p=1}^{P}\lambda_p\boldsymbol{\rho^{(p)}}=\boldsymbol{\rho}\end{aligned} \tag{1.1}$$

1.3.1 Optimality of full disclosure

For any given belief $\tilde{\boldsymbol{\rho}}$, let $\mathrm{DV}(\tilde{\boldsymbol{\rho}})$ denote the difference between the expected valuations of types 1 and 2 under $\tilde{\boldsymbol{\rho}}$, i.e., $\mathrm{DV}(\tilde{\boldsymbol{\rho}}) \doteq \langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{1\cdot}}\rangle - \langle\tilde{\boldsymbol{\rho}},\boldsymbol{v_{2\cdot}}\rangle$. Our first result shows that full disclosure is optimal if and only if the underlying uncertainty cannot affect the ranking of the types.

Proposition 1.1 (Optimality of full disclosure)

Full disclosure is optimal for the seller if and only if the types are ranked uniformly across states: either $v_{1k} \geqq v_{2k}$ holds for all k or $v_{2k} \geqq v_{1k}$ holds for all k.

Proof.

Sufficiency: Assume that $v_{1k} \geqq v_{2k}$ holds for every k. Under this assumption, type 1's expected valuation is at least as high as type 2's, no matter what the underlying belief is.

Consider now an arbitrary distribution of posteriors $(\boldsymbol{\rho^{(1)}}, \ldots, \boldsymbol{\rho^{(P)}}; \boldsymbol{\lambda})$ that is consistent with the law of total probability. At every posterior belief $\boldsymbol{\rho^{(p)}}$, it is optimal for the seller either to charge type 1 what he is willing to pay and sell only to him, or to sell to both types at type 2's expected valuation. Now replace the belief $\boldsymbol{\rho^{(p)}}$ by its decomposition into a distribution of fully revealing posteriors. The seller can always set the price such that the type that was marginal at $\boldsymbol{\rho^{(p)}}$ remains marginal at every fully revealing posterior. This strategy keeps the expected price the same. Since the ranking of the types is the same across states, the types that purchased the good at $\boldsymbol{\rho^{(p)}}$ are still buying, which weakly increases the expected profit:

$$\Pi(\boldsymbol{\rho^{(p)}}; \alpha) \leqq \sum_{k=1}^{K} \rho_k^{(p)} \Pi(\boldsymbol{e^{(k)}}; \alpha).$$

The same argument applies to every posterior in $(\boldsymbol{\rho^{(1)}}, \ldots, \boldsymbol{\rho^{(P)}}; \boldsymbol{\lambda})$. Therefore, the seller can always achieve a weakly higher profit using fully revealing posteriors $(\boldsymbol{e^{(1)}}, \ldots, \boldsymbol{e^{(K)}}; \boldsymbol{\rho})$:

$$\sum_{p=1}^{P} \lambda_p \Pi(\boldsymbol{\rho^{(p)}}; \alpha) \leqq \sum_{p=1}^{P} \lambda_p \sum_{k=1}^{K} \rho_k^{(p)} \Pi(\boldsymbol{e^{(k)}}; \alpha) = \sum_{k=1}^{K} \left(\sum_{p=1}^{P} \lambda_p \rho_k^{(p)} \right) \Pi(\boldsymbol{e^{(k)}}; \alpha) = \sum_{k=1}^{K} \rho_k \Pi(\boldsymbol{e^{(k)}}; \alpha).$$

Since this is true for any distribution of posterior beliefs consistent with the law of total probability, it follows that fully disclosing the state must be at least weakly optimal for the seller.

Necessity: If types are not ranked uniformly for every state, there exist states ψ_{k_1} and ψ_{k_2} such that $\mathrm{DV}(\boldsymbol{e^{(k_1)}}) > 0$ and $\mathrm{DV}(\boldsymbol{e^{(k_2)}}) < 0$. Since the function DV is linear, there exists a unique $\gamma^* \in (0, 1)$ such that the posterior belief $\boldsymbol{\rho^*} \doteq \gamma^* \boldsymbol{e^{(k_1)}} + (1 - \gamma^*) \boldsymbol{e^{(k_2)}}$ equalizes the expected valuations of the two types, i.e., $\mathrm{DV}(\boldsymbol{\rho^*}) = 0$.

We show that disclosing no further information at $\boldsymbol{\rho}^*$ is strictly better than decomposing $\boldsymbol{\rho}^*$ into a distribution of fully revealing beliefs. Notice that at $\boldsymbol{\rho}^*$, expected valuations are equalized. Therefore, the seller can charge this expected valuation and sell to both types without leaving them any surplus. If the actual state is fully disclosed, however, the seller loses profit on at least one of the types. First, if the state is known, types are not willing to pay more than their valuation in that state. Therefore, the maximal expected profit obtainable from a type under full disclosure is bounded from above by the original expected valuation. Second, we know that in state ψ_{k_1}, the inequality $v_{1k_1} > v_{2k_1}$ is true. Therefore, if this state is revealed, either type 1 does not pay his entire valuation or type 2 does not buy the good under full disclosure. In either case, the seller loses expected profit.

Now, return to the prior $\boldsymbol{\rho}$, and consider the distribution of fully revealing posteriors that leads to belief $\boldsymbol{e}^{(\boldsymbol{k})}$ with probability ρ_k for each k. Since $\boldsymbol{\rho} \in \operatorname{int} \Delta^{K-1}$ by assumption, it follows that $\rho_k > 0$ for every k. Therefore, there exists an $\varepsilon > 0$ small enough such that $\varepsilon\gamma^* \leqq \rho_{k_1}$ and $\varepsilon(1-\gamma^*) \leqq \rho_{k_2}$ are both true. Using this ε, we can reduce the probabilities assigned to posteriors $\boldsymbol{e}^{(\boldsymbol{k_1})}$ and $\boldsymbol{e}^{(\boldsymbol{k_1})}$ by $\varepsilon\gamma^*$ and $\varepsilon(1-\gamma^*)$, respectively, and assign probability ε to posterior $\boldsymbol{\rho}^*$. By the definition of $\boldsymbol{\rho}^*$, the new distribution still satisfies the law of total probability. Since no disclosure at $\boldsymbol{\rho}^*$ gives a strictly higher profit than its decomposition into a distribution of fully revealing beliefs, the modified distribution leads to a strictly higher expected profit for the seller than does full disclosure. ■

1.3.2 Structure of the optimal signal

Note that the result of the proposition is quite strong since it does not depend on either the prior beliefs $\boldsymbol{\rho}$ or the distribution of the buyer's type $\boldsymbol{\alpha}$. The proof shows that if the two types are not ranked uniformly, then full disclosure is never optimal since the seller always prefers to reallocate some positive probability to a posterior at which both types have the same expected valuation. In fact, this observation can be used to describe the structure of the optimal posteriors more closely.

In what follows, we assume without loss of generality that type 1 has a weakly higher valuation under the prior distribution, i.e., $\mathrm{DV}(\boldsymbol{\rho}) \geqq 0$. To simplify the exposition, we assume from now on that types are strictly ranked in every state, i.e., $v_{1k} \neq v_{2k}$ for every k. We partition the set of states into two subsets according to the ranking of the types in terms of their valuations: $\Psi = \Psi^{(1)} \cup \Psi^{(2)}$, where $\Psi^{(i)} \doteq \{\psi_k \in \Psi \colon v_{ik} > v_{-ik}\}$ for each $i = 1, 2$. Moreover, we assume without loss of generality that states are labeled in such a way that $\Psi^{(1)}$ contains the first J states ($0 \leqq J \leqq K$) and $\Psi^{(2)}$ the remaining $K - J$ states. For each $\psi_{k_1} \in \Psi^{(1)}$ and each $\psi_{k_2} \in \Psi^{(2)}$, let $\boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_2})}$ denote the unique convex combination of $\boldsymbol{e}^{(\boldsymbol{k_1})}$ and $\boldsymbol{e}^{(\boldsymbol{k_2})}$ that equalizes the expected valuations of the two types. Formally,

$$\boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_2})} \doteq \frac{v_{2k_2} - v_{1k_2}}{v_{1k_1} - v_{2k_1} + v_{2k_2} - v_{1k_2}} \boldsymbol{e}^{(\boldsymbol{k_1})} + \frac{v_{1k_1} - v_{2k_1}}{v_{1k_1} - v_{2k_1} + v_{2k_2} - v_{1k_2}} \boldsymbol{e}^{(\boldsymbol{k_2})}.$$

Proposition 1.2 (support of the optimal distribution) Assume $\mathrm{DV}(\boldsymbol{\rho}) \geqq 0$. There exists an optimal distribution of posterior beliefs such that its support is contained in

$$\left\{\boldsymbol{e}^{(\boldsymbol{k})} \colon \psi_k \in \Psi^{(1)}\right\} \bigcup \left\{\boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_2})} \colon \psi_{k_1} \in \Psi^{(1)}, \psi_{k_2} \in \Psi^{(2)}\right\}.$$

Proof. First, note that the seller never wants to assign positive probability to two posterior beliefs that induce the opposite ranking of the two types. If this is not the case, using the logic described in the proof of the necessity part of Proposition 1.1, there is a convex combination of two such posteriors that equalizes the expected valuations of the two types and captures more expected surplus. Such a convex combination is strictly preferred to revealing both posteriors with positive probability. Since the optimal distribution of posteriors has to satisfy the law of total probability, we can restrict attention to the set $H_+ \doteq \left\{\tilde{\boldsymbol{\rho}} \in \Delta^{K-1} \colon \mathrm{DV}(\tilde{\boldsymbol{\rho}}) \geqq 0\right\}$. The set H_+ is a convex polytope contained in the $K-1$-dimensional simplex.

In a way similar to the proof of the sufficiency part of Proposition 1.1, we can show that every posterior belief $\tilde{\boldsymbol{\rho}} \in H_+$ can be replaced by its decomposition into a distribution of the extreme points of the polytope H_+ without losing any expected profit. This is true since every such posterior belief can be expressed as a convex combination of the finitely many extreme beliefs, and such beliefs never reverse the ranking of the types compared to $\tilde{\boldsymbol{\rho}}$.

Hence, we only need to show that the extreme points of H_+ are exactly the beliefs listed in the statement of the proposition. First, note that the extreme points of the simplex Δ^{K-1} are exactly the fully revealing beliefs $\left\{\boldsymbol{e^{(k)}} \colon k = 1, \ldots, K\right\}$. Since H_+ is the intersection of the simplex and the half-space defined by the linear inequality $\mathrm{DV}(\tilde{\boldsymbol{\rho}}) \geqq 0$, all fully revealing beliefs that rank type 1 above type 2 are still contained in this set. Therefore, these beliefs must be extreme points of the set H_+ as well. The intersection with the half-space $\{\tilde{\boldsymbol{\rho}} \colon \mathrm{DV}(\tilde{\boldsymbol{\rho}}) \geqq 0\}$ might also introduce new extreme points at which the additional linear inequality constraint $\mathrm{DV}(\tilde{\boldsymbol{\rho}}) \geqq 0$ is binding. This means that, to complete the set of extreme points of H_+, we only need to find the extreme points of the face $H_0 \doteq \Delta^{K-1} \bigcap \{\tilde{\boldsymbol{\rho}} \colon \mathrm{DV}(\tilde{\boldsymbol{\rho}}) = 0\} \subseteq H_+$.

First, for each $\psi_{k_1} \in \Psi^{(1)}$ and each $\psi_{k_2} \in \Psi^{(2)}$, the vector $\boldsymbol{b^{(k_1,k_2)}}$ is an extreme point of the set H_0. To see this, take two vectors $\boldsymbol{y^{(1)}}, \boldsymbol{y^{(2)}} \in H_0$ and a scalar $\gamma \in (0,1)$ such that $\boldsymbol{b^{(k_1,k_2)}} = \gamma \boldsymbol{y^{(1)}} + (1-\gamma)\boldsymbol{y^{(2)}}$. Since $\boldsymbol{b^{(k_1,k_2)}}$ assigns probability 0 to every state other than ψ_{k_1} and ψ_{k_2}, the vectors $\boldsymbol{y^{(1)}}$ and $\boldsymbol{y^{(2)}}$ must do the same. This observation together with $\mathrm{DV}(\boldsymbol{y^{(1)}}) = \mathrm{DV}(\boldsymbol{y^{(2)}}) = 0$ imply that both vectors must assign positive probability to both states ψ_{k_1} and ψ_{k_2}. We know that there is a unique vector with these properties, namely, $\boldsymbol{b^{(k_1,k_2)}} = \boldsymbol{y^{(1)}} = \boldsymbol{y^{(2)}}$.

Second, we show that every extreme point of H_0 is of the form $\boldsymbol{b^{(k_1,k_2)}}$ for some $\psi_{k_1} \in \Psi^{(1)}$ and $\psi_{k_2} \in \Psi^{(2)}$. It follows immediately that the support of an extreme point of H_0 cannot contain a single state (since, by assumption, valuations are not equal in any state), and all extreme points that assign positive probability to exactly two states are of the form $\boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_2})}$. Therefore, consider a vector $\boldsymbol{b} = \sum_{k=1}^{K} \beta_k \boldsymbol{e}^{(\boldsymbol{k})} \in H_0$, such that $\beta_k \geqq 0$ for every k, and $\sum_{k=1}^{K} \beta_k = 1$, and $\boldsymbol{b}$ assigns positive probability to at least 3 states. Assume that $\beta_{k_1}, \beta_{k_2}, \beta_{k_3} > 0$, where $\psi_{k_1} \in \Psi^{(1)}$ and $\psi_{k_2}, \psi_{k_3} \in \Psi^{(2)}$. Define $\boldsymbol{d} \doteq \boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_2})} - \boldsymbol{b}^{(\boldsymbol{k_1},\boldsymbol{k_3})}$, and take the vectors $\boldsymbol{b} + \varepsilon \boldsymbol{d}$ and $\boldsymbol{b} - \varepsilon \boldsymbol{d}$. For $\varepsilon > 0$ small enough, $\boldsymbol{b} + \varepsilon \boldsymbol{d}, \boldsymbol{b} - \varepsilon \boldsymbol{d} \in \Delta^{K-1}$, and by the linearity of DV, $\boldsymbol{b} + \varepsilon \boldsymbol{d}, \boldsymbol{b} - \varepsilon \boldsymbol{d} \in H_0$. Since $\boldsymbol{b} = \frac{1}{2}(\boldsymbol{b} + \varepsilon \boldsymbol{d}) + \frac{1}{2}(\boldsymbol{b} - \varepsilon \boldsymbol{d})$, and $\boldsymbol{d} \neq \boldsymbol{0}$, the vector $\boldsymbol{b}$ cannot be an extreme point of H_0. Analogous steps show the same for the case when $\psi_{k_1}, \psi_{k_2} \in \Psi^{(1)}$ and $\psi_{k_3} \in \Psi^{(2)}$, which completes the proof. ■

Again, note that the result does not depend on the distribution of the buyer's types $\boldsymbol{\alpha}$, and is true whether or not the valuations of the two types are ranked uniformly across states.

The optimal information policy always induces the same weak ranking as the prior belief and helps the seller increase expected profit by (i) conditioning the group of types to which she sells the product on the released information, and (ii) introducing posterior beliefs that equalize the expected valuations to capture more surplus. It is important not to interpret the result as the optimality of "full revelation" over a subpolytope defined by the function DV. Although it is optimal to use only the extreme points of this subpolytope, such a decomposition is not unique if the ranking is not uniform over states (and hence the subpolytope is a strict subset of the probability simplex). The following proposition describes the optimal distribution of the extreme points and shows that the seller has an incentive to fully reveal states in which the net gain from selling only to the higher valuation type (over selling to both types) is above a certain threshold.

Proposition 1.3 If $\boldsymbol{\lambda}^* \in \mathbb{R}^J$ and $\boldsymbol{\eta}^* \in \mathbb{R}^{J \times (K-J)}$ constitute an optimal distribution over the set of posteriors described in Proposition 1.2, then there exists a threshold $T > 0$ such that for every $k \leqq J$,

$$\lambda_k^* = \begin{cases} \rho_k & \text{if } \alpha v_{1k} - v_{2k} > T; \\ 0 & \text{if } 0 < \alpha v_{1k} - v_{2k} < T. \end{cases} \tag{1.2}$$

Moreover, if $\lambda_k^* < \rho_k$ for some k such that $\alpha v_{1k} - v_{2k} > 0$, then $\lambda_l = 0$ must hold for every $l \leqq J$ such that $\alpha v_{1l} - v_{2l} \leqq 0$.

Proof. See Appendix A.1. ■

Alonso and Câmara (2016) derive a similar cutoff structure in the context of persuading voters to approve a proposal with an uncertain outcome. In the special case of a single voter, they show that optimally, the proposal is accepted in states in which the receiver's (i.e., the voter's) net gain is above a certain cutoff and rejected otherwise. The intuition behind their result is that the sender (i.e., the politician) can decrease the "average" probability of rejection by pooling states in which the voter would approve the proposal with some states in which he would reject it, such that the final decision is unchanged. They prove that the optimal way of doing this is by adding states of the second kind for which the voter's incentive to reject the proposal is the smallest until the voter is indifferent between accepting and rejecting.

Our result provides a slightly different intuition in a more complicated setting. The sender's (i.e., the seller's) objective is more complex since she has to consider both the probability of sale and the price that can be charged for the product. Therefore, her payoff is a continuous, piecewise linear function of the posterior belief which typically possesses both concave and convex features instead of a discontinuous, piecewise constant function reflecting the payoff from accepting or rejecting a proposal. This difference between the settings of the two papers leads to a different linear programming problem and different features of the cutoff property. If we think about the two alternatives in our model as "charging the higher type's valuation" and "charging the lower type's valuation," then it is true that the seller will optimally choose the first one in states in which her net gain is above the threshold and the second one otherwise. This is driven purely by the seller's incentive to maximize the expected net gain from restricting output only for states in which this net gain is positive. However, contrary to the model in Alonso and Câmara (2016), in our model she will not necessarily be indifferent between the two alternatives if she chooses the lower type's valuation since some fully revealing posteriors for which the net gain is negative but the ranking is the same as under the prior might also receive positive probability in the optimal solution. Moreover, the role of the two alternatives and the order of the different states in the cutoff structure are completely reversed for prior beliefs that induce the opposite ranking.

The following example illustrates the threshold structure of the optimal solution in a simple setting.

Example 1.1 Consider the four-state environment with valuations as defined in Table 1.1. Assume that the prior belief is given by $\boldsymbol{\rho} = (0.3, 0.3, 0.2, 0.2)$, and that both types are equally likely, i.e., $\alpha = 1/2$.

v	ψ_1	ψ_2	ψ_3	ψ_4
θ_1	6	5	4	2
θ_2	2	2	3	4

Table 1.1: Valuations in Example 1.1

Type 1 has a strictly higher valuation in the first three states, while this ranking is reversed in the last state. The expected valuations of the types under the prior belief are $\langle \boldsymbol{v_{1\cdot}}, \boldsymbol{\rho} \rangle = 4.5$ and $\langle \boldsymbol{v_{2\cdot}}, \boldsymbol{\rho} \rangle = 2.6$. Therefore, by Proposition 1.2, the optimal disclosure policy can be constructed using (i) some fully revealing posteriors that rank type 1 above type 2, and (ii) beliefs that equalize the expected valuations of the two types by concealing information between two states that lead to opposite rankings. These posteriors are listed in the first row of Table 1.2.

$(1,0,0,0)$	$(0,1,0,0)$	$(0,0,1,0)$	$(\frac{1}{3},0,0,\frac{2}{3})$	$(0,\frac{2}{5},0,\frac{3}{5})$	$(0,0,\frac{2}{3},\frac{1}{3})$
0.3	0.233	0	0	0.167	0.3

Table 1.2: Optimal disclosure policy in Example 1.1

The second row of the table contains the optimal probabilities obtained by solving the corresponding optimization problem. The optimal solution satisfies the threshold structure described in Proposition 1.3. The net gain from restricting the output and selling only to type 1 is positive for the first two states, and is the highest in the first state. The optimal solution reveals state ψ_1 with probability 1 and state ψ_2 with probability $0.233/0.3 = 0.778$. Since not all states with a positive net gain are revealed with probability 1, the cutoff structure generalizes to state 3 as well. □

We can summarize the findings of Propositions 1.1-1.3 about the optimal disclosure policy as follows. If types are ranked in the same way in every state, then a fully revealing distribution of posteriors is optimal. The seller strictly benefits from disclosing information if and only if the optimal groups of types purchasing the good are not the same in every state. If types are not ranked uniformly, then full disclosure is never optimal. An optimal, partially disclosing distribution of posteriors can be found by solving a linear programming problem described in the proof of Proposition 1.3. The optimal signal never reverses the ranking of the types according to their valuation. Moreover, the seller prefers to restrict output and sell only to the higher type in states in which the net profit from this is strictly above a certain threshold, and prefers to sell to both types in states in which the net utility is strictly below this threshold. The seller strictly benefits from information disclosure if and only if $\mathrm{DV}(\boldsymbol{\rho}) > 0$ and there is at least one state in which it is optimal to sell only to the higher type.

1.3.3 Horizontal-vertical decomposition of the state space

Management science papers on optimal consumer feedback typically model product-related uncertainty by assuming a two-dimensional structure. A vertical product characteristic (e.g., build quality) is perceived by all buyer types in the same way and typically changes the buyers' willingness to pay by the same amount. A horizontal characteristic (e.g., fitness of a product for a particular task a consumer has in mind) has heterogeneous effects on the valuations of different types. These papers typically analyze the effects of information disclosure on different market settings in the presence of horizontal and vertical product attributes. Jiang and Guo (2015) consider the restricted information-design problem of a monopolistic seller in an environment in which buyers report their post-consumption utility on a scale whose coarseness is set by the seller. They show that if vertical attributes are relatively more important than horizontal attributes, then it is optimal for the seller to provide as much information as possible and she does this by picking the finest scale. On the other hand, if horizontal characteristics are important enough, the seller's optimal strategy is to reveal as little as possible which can be implemented by choosing the coarsest scale. In this section, we show that the two-type model can be easily transformed into a setting with a two-dimensional state space in which the horizontal and vertical dimensions of information are measured by the different state variables. Our two-dimensional decomposition highlights the connection to this literature, and provides additional theoretical insights to its results.

The horizontal and vertical dimensions in our model are based on the following observations. First, by the definition of the sets $\Psi^{(1)}$ and $\Psi^{(2)}$, knowing which one of these sets the true state ψ belongs to completely determines the order of the types according to their willingness to pay. Therefore, we can define a binary state variable $\xi \in \Xi \doteq \{\xi_1, \xi_2\}$ that captures this *horizontal dimension* of information: $\xi = \xi_i$ if and only if $\psi \in \Psi^{(i)}$ for each $i = 1, 2$.

Second, conditional on the realization of the horizontal state variable ξ, the remaining uncertainty cannot change the ranking of the types anymore. Therefore, we can define two separate state variables that describe this *vertical dimension* of information as follows. If type 1 has a strictly higher valuation than type 2 (i.e., $\psi \in \Psi^{(1)}$ and $\xi = \xi_1$), then the realization of the first vertical state variable $\omega^{(1)} \in \Omega^{(1)} \doteq \left\{\omega_1^{(1)}, \ldots, \omega_J^{(1)}\right\}$ pins down the actual state: if $\psi = \psi_k$, where $k \leqq J$, then $\omega^{(1)} = \omega_k^{(1)}$ must be true. Similarly, if type 2 has a higher valuation in state ψ, then the second vertical state variable $\omega^{(2)} \in \Omega^{(2)} \doteq \left\{\omega_{J+1}^{(2)}, \ldots, \omega_K^{(2)}\right\}$ determines the true state: if $\psi = \psi_k$, where $J \leqq k \leqq K$, then $\omega^{(2)} = \omega_k^{(2)}$ must hold.

Notice that every state $\psi \in \Psi$ restricts the value of the single vertical state variable that corresponds to the ranking of the types under ψ. The value of the other vertical variable can be arbitrary. Therefore, for every state in the original state space Ψ, there might be multiple states in the decomposed state space $\Xi \times \Omega^{(1)} \times \Omega^{(2)}$ that represent the same preferences. This also implies that there might be multiple ways of defining the underlying prior distribution on $\Xi \times \Omega^{(1)} \times \Omega^{(2)}$ that are consistent with the prior $\boldsymbol{\rho}$ on Ψ.[6] Notice also that we can merge the two vertical dimensions into one by using $\Omega \doteq \Omega^{(1)} \times \Omega^{(2)}$ such that the new vertical state variable $\omega \in \Omega$ still cannot alter the ranking of the types if the horizontal state ξ is known.

This distinction between the horizontal and vertical dimensions is less restrictive than the definition mentioned at the beginning of the section. Importantly, it is still possible for the vertical dimension of information to have some heterogeneous effects on the types. Moreover, the valuations of the different types do not even have to move in the same direction if the vertical state changes; the only requirement is that their *ranking* has to remain the same.

[6]Essentially, we split up every state in Ψ by the realization of the irrelevant vertical state variable and organize the resulting set of states in a multidimensional array.

The next example illustrates the horizontal-vertical decomposition in a simple four-state environment.

Example 1.2 Consider the valuations given in Table 1.3.

v	ψ_1	ψ_2	ψ_3	ψ_4
θ_1	5	4	2	1
θ_2	2	3	4	4

Table 1.3: Valuations in Example 1.2

Type 1 has a strictly higher valuation than type 2 in the first two states, and this ranking is reversed in the last two states. Therefore, the horizontal state variable ξ selects whether the actual state ψ belongs to the first two states ($\xi = \xi_1$) or the last two states ($\xi = \xi_2$). Given the realization of the horizontal state variable, $\xi = \xi_i$, the vertical state variable $\omega^{(i)}$ determines the actual state. For example, if the state is $\psi = \psi_2$, then $\xi = \xi_1$ and $\omega^{(1)} = \omega_2^{(1)}$ while $\omega^{(2)}$ can be arbitrary. The valuation profiles of the types in the decomposed state space are illustrated in Table 1.4, where the rows and columns represent the different possible realizations of the horizontal state and the vertical state, respectively.

	$(\omega_1^{(1)}, \omega_3^{(2)})$	$(\omega_1^{(1)}, \omega_4^{(2)})$	$(\omega_2^{(1)}, \omega_3^{(2)})$	$(\omega_2^{(1)}, \omega_4^{(2)})$
ξ_1	$(5, 2)$	$(5, 2)$	$(4, 3)$	$(4, 3)$
ξ_2	$(2, 4)$	$(1, 4)$	$(2, 4)$	$(1, 4)$

Table 1.4: Two-dimensional decomposition: valuations

A belief $\boldsymbol{\nu}$ on the decomposed state space is simply a three-dimensional array of numbers such that ν_{ijk} denotes the probability that (i) the horizontal state is ξ_i and (ii) the vertical states are $(\omega_j^{(1)}, \omega_k^{(2)})$, for each $i = 1, 2$, $j = 1, 2$, $k = 3, 4$, where $\nu_{i,j,k} \geqq 0$ for each i, j, k, and $\sum_{i,j,k} \nu_{ijk} = 1$. This distribution is illustrated in Table 1.5.

	$(\omega_1^{(1)}, \omega_3^{(2)})$	$(\omega_1^{(1)}, \omega_4^{(2)})$	$(\omega_2^{(1)}, \omega_3^{(2)})$	$(\omega_2^{(1)}, \omega_4^{(2)})$
ξ_1	ν_{113}	ν_{114}	ν_{123}	ν_{124}
ξ_2	ν_{213}	ν_{214}	ν_{223}	ν_{224}

Table 1.5: Two-dimensional decomposition: beliefs

For a given belief $\tilde{\boldsymbol{\rho}}$ on the original state space Ψ, there are multiple ways of defining an equivalent belief $\tilde{\boldsymbol{\nu}}$ on the decomposed state space $\Xi \times \Omega$. Using the previously defined notation, the belief $\tilde{\boldsymbol{\nu}}$ has to satisfy the following system of equations:

$$\tilde{\rho}_1 = \tilde{\nu}_{113} + \tilde{\nu}_{114}; \qquad \tilde{\rho}_2 = \tilde{\nu}_{123} + \tilde{\nu}_{124};$$

$$\tilde{\rho}_3 = \tilde{\nu}_{213} + \tilde{\nu}_{223}; \qquad \tilde{\rho}_4 = \tilde{\nu}_{214} + \tilde{\nu}_{224}.$$

□

As we saw above, the horizontal dimension can change the ranking of the types, while the vertical dimension has no effect on the order once the horizontal state is known. Given this horizontal-vertical decomposition, Proposition 1.2 can be restated as follows.

Corollary 1.1 The state space Ψ can be decomposed into a horizontal-vertical state space $\Xi \times \Omega$ such that there is an optimal distribution of posteriors in which every posterior either

- fully reveals both the horizontal and the vertical state such that the induced ranking is the same as under the prior, or
- combines the two horizontal states but fully reveals the vertical state such that expected valuations are equalized.

Proof. See Appendix A.2. ■

Corollary 1.1 shows that in a less restrictive information-design environment than the one analyzed in Jiang and Guo (2015) and with more general definitions of the horizontal and vertical attributes, it is optimal for the seller to disclose all information about the vertical dimension Ω; in addition, if both rankings can occur with positive probability, information about the horizontal component should be partially disclosed. For the seller to achieve the highest level of expected profit, disclosure typically does not happen independently along the two dimensions.

Example 1.3 Continue Example 1.2 and assume that the prior belief is given by the vector $\boldsymbol{\rho} = (0.4, 0.2, 0.3, 0.1)$, and that both types are equally likely, i.e., $\alpha = 0.5$. Since type 1 has a higher expected valuation at the prior, $\langle \boldsymbol{\rho}, \boldsymbol{v_{1\cdot}} \rangle = 3.5$ and $\langle \boldsymbol{\rho}, \boldsymbol{v_{2\cdot}} \rangle = 3$, by Proposition 1.2 the optimal posterior distribution can be constructed using the posteriors contained in the first column of Table 1.6.

posteriors on Ψ	posteriors on $\Xi \times \Omega^{(1)} \times \Omega^{(2)}$	optimal weights
$(1, 0, 0, 0)$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	1/6
$(0, 1, 0, 0)$	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	0
$(2/5, 0, 3/5, 0)$	$\begin{bmatrix} 2/5 & 0 & 0 & 0 \\ 3/5 & 0 & 0 & 0 \end{bmatrix}$	1/2
$(1/2, 0, 0, 1/2)$	$\begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{bmatrix}$	1/15
$(0, 2/3, 1/3, 0)$	$\begin{bmatrix} 0 & 0 & 2/3 & 0 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}$	0
$(0, 3/4, 0, 1/4)$	$\begin{bmatrix} 0 & 0 & 0 & 3/4 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}$	4/15

Table 1.6: Optimal distribution of posteriors

The second column of the table presents the horizontal-vertical equivalents of these posteriors. In the table, we can see that all of these decomposed posterior beliefs satisfy the properties described in Corollary 1.1. They disclose all information along the vertical dimension while they only partially disclose information along the horizontal dimension. The optimal weights are listed in the third column of the table. □

1.3.4 Welfare effects of the optimal disclosure policy

The last result of this section describes the effects of optimal disclosure on welfare. Since the optimal information structure helps the seller to condition the group of types that buy the product on the released information without reversing the ranking of the types, the effects on welfare will be related to the change in the probability of a transaction.

Proposition 1.4 (Welfare effects) Let FD, ND, and $*$ denote full disclosure, no disclosure, and the optimal disclosure policy, respectively. Then the relationship between the total surplus, the producer surplus, and the consumer surplus under the three disclosure policies can be described by the following figure, where the symbol $\lesseqqgtr$ means that the relationship can go either way depending on the model parameters.

$$
\begin{array}{ccc|ccc|ccc}
 & \mathrm{TS}^{*} & & & \mathrm{PS}^{*} & & & \mathrm{CS}^{*} & \\
\rotatebox{-45}{$\lesseqqgtr$} & & \rotatebox{45}{$\leqq$} & \rotatebox{-45}{$\leqq$} & & \rotatebox{45}{$\leqq$} & \rotatebox{-45}{$\lesseqqgtr$} & & \rotatebox{45}{$\geqq$} \\
\mathrm{TS}^{\mathrm{ND}} & \lesseqqgtr & \mathrm{TS}^{\mathrm{FD}} & \mathrm{PS}^{\mathrm{ND}} & \lesseqqgtr & \mathrm{PS}^{\mathrm{FD}} & \mathrm{CS}^{\mathrm{ND}} & \lesseqqgtr & \mathrm{CS}^{\mathrm{FD}}
\end{array}
$$

Proof. We will use the following two examples throughout the proof to establish that some comparisons can go either way. In the examples, there are two equally likely states and two equally likely types. The valuations are given in Tables 1.7 (a) and 1.7 (b), respectively. The expected valuations of the two types and the indirect profit as a function of ρ_1 are illustrated in Figures 1.1 (a) and 1.1 (b). In these figures, the solid gray lines represent the expected valuations of the two types, the dotted lines are the expected valuations multiplied by the mass of the type, and the thick black lines illustrate the indirect profit. For the first example, some expected valuations, optimal prices, and indirect profit levels are also computed in Table 1.8.

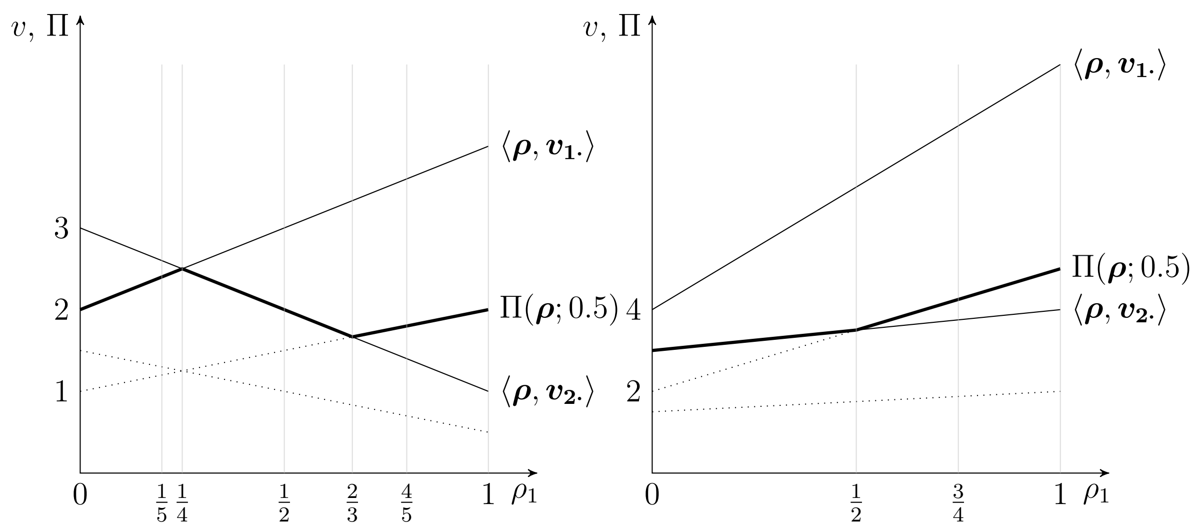
v	ψ_1	ψ_2
θ_1	4	2
θ_2	1	3

(a) first example

v	ψ_1	ψ_2
θ_1	10	4
θ_2	4	3

(b) second example

Table 1.7: Valuations in the examples



(a) first example (b) second example

Figure 1.1: Expected valuations and indirect profit as a function of ρ_1

Total surplus Since the seller faces no production costs, it is always efficient to sell the good to both types. Therefore, the total surplus under a disclosure policy is related to the probability of sale.

ρ_1	valuation of type 1	valuation of type 2	optimal price	indirect profit
0	2	3	2	2
1/5	2.4	2.6	2.4	2.4
1/4	2.5	2.5	2.5	2.5
1/2	3	2	2	2
2/3	3.33	1.67	1.67	1.67
4/5	3.6	1.4	3.6	1.8
1	4	1	4	2

Table 1.8: Sample values for the first example

1. $\mathrm{TS}^* \geqq \mathrm{TS}^{FD}$: The difference between the optimal policy and full disclosure is that in the former, probability from some pairs of fully revealing posteriors that lead to different rankings are shifted to posteriors that equalize the expected valuations of the two types. Since both types trivially purchase the product at the latter posteriors, the product is sold with a higher probability than under full disclosure, increasing total surplus.

2. $\mathrm{TS}^{ND} \lesseqqgtr \mathrm{TS}^{FD}$: Consider the first example. If the prior is $\rho_1 = 4/5$ and no further information is disclosed, only the first type purchases the product and the probability of sale is 0.5. On the other hand, the probability of sale is strictly higher if the state is fully revealed since both types buy the product in state 2 (i.e., at posterior 0).

   For the other direction, consider the prior $\rho_1 = 1/2$. Moving from no disclosure to full disclosure decreases the probability of sale from 1 to some strictly lower number since only type 1 buys the product if state 1 is revealed (i.e., at posterior 1).

3. $\mathrm{TS}^{ND} \lesseqqgtr \mathrm{TS}^*$: Once again, take the first example and consider priors 4/5 and 1/2. In both cases, the optimal policy assigns positive probabilities to two posteriors: 1/4 and 1. At posterior 1/4, both types purchase the product, and only type 1 buys it a

posterior 1. Therefore, moving from no disclosure to the optimal policy increases the probability of sale for the first prior, and decreases this probability for the second one.

Producer surplus

1. $\mathrm{PS}^* \geqq \mathrm{PS}^{FD}$ and $\mathrm{PS}^* \geqq \mathrm{PS}^{ND}$ are true since both full disclosure and no disclosure are feasible policies in the seller's optimal information disclosure problem.

2. $\mathrm{PS}^{ND} \lesseqqgtr \mathrm{PS}^{FD}$: Consider the first example. Since the indirect profit is 2 for posteriors 0 and 1, the profit from full disclosure is also equal to 2 regardless of the prior beliefs. The optimal profit from no disclosure is larger than this level for the prior $\rho_1 = 1/4$ (2.5), and smaller than this for $\rho_1 = 4/5$ (1.8).

Consumer surplus

1. $\mathrm{CS}^* \leqq \mathrm{CS}^{FD}$: Once again, we can move from full disclosure to the optimal policy by concealing some information between some pairs of states that lead to different rankings. At these newly introduced posteriors, the two types have the same valuation, and all the surplus is extracted by the seller. Therefore, moving from full disclosure to the optimal policy weakly reduces consumer surplus.

2. $\mathrm{CS}^{ND} \lesseqqgtr \mathrm{CS}^{FD}$: Take the first example, and consider prior $\rho_1 = 4/5$. Under no disclosure, only type 1 buys the product, and consumer surplus is 0. Full disclosure reveals state 2 with positive probability, in which case both types purchase the good at price 2, leaving some positive consumer surplus for the second type.

   For the other direction, take prior $\rho_1 = 1/2$. If no information is disclosed, both types buy the product at price 2, and the consumer surplus is $1/2 \cdot 1 = 1/2$. Full disclosure is equivalent to a 50-50 mixture between fully revealing state 1 and fully revealing state 2, leading to an expected surplus of $1/2 \cdot 0 + 1/2 \cdot 1/2 \cdot 1 = 1/4$.

3. $\text{CS}^{ND} \lesseqqgtr \text{CS}^*$: In the first example, if no information is disclosed at the prior belief $\rho_1 = 1/2$, both types buy the product and there is positive consumer surplus. The optimal policy is a lottery over posteriors 1/4 and 1, where both of these beliefs lead to 0 consumer surplus. Therefore, no disclosure is strictly better for the consumers.

   For the other direction, consider the second example. Here, full disclosure is always optimal. If the prior belief is $\rho_1 = 3/4$ and no information is disclosed, only type 1 buys the product, and consumer surplus is zero. On the other hand, the optimal policy places positive probability on fully revealing state 2 (i.e., on posterior 0), in which case consumer surplus is positive. ■

The most interesting observation here is that the optimal disclosure policy is more efficient than full disclosure (i.e., $\text{TS}^* \geqq \text{TS}^{FD}$), but more than this efficiency gain is extracted by the seller (since $\text{CS}^* \leqq \text{CS}^{FD}$).

We have to be careful when interpreting this result. Depending on the particular context, reaching the above information structure might have other associated welfare effects. For example, in the case of online stores with consumer review systems, the information about the product is most likely revealed gradually over time by its previous consumers. If the amount of information revealed is correlated with the number of items sold, the seller might want to adjust the speed of information disclosure by lowering the prices in earlier periods. This can lead to additional welfare effects that should also be taken into account.

1.4 Generalization to many types

In this section, we present the generalizations of some of the above results to the many-type environment. We derive results analogous to Propositions 1.1 and 1.2 and present necessary properties that the optimal distribution of posteriors must satisfy. With more than two types, it is still true that the horizontal dimension of information gives the seller incentives to disclose less information, but these may not necessarily be strong enough to render full disclosure suboptimal. In general, the optimality of full disclosure depends on the underlying type distribution $\boldsymbol{\alpha}$.

To set up the seller's many-type disclosure problem, consider the environment with I types, i.e., $\Theta = \{\theta_1, \ldots, \theta_I\}$. For any subset of types $B \subseteq \Theta$, let $\alpha(B) \doteq \sum_{\theta_i \in B} \alpha_i$ denote the mass of subset B. For any $B \subseteq \Theta$ and $\tilde{\boldsymbol{\rho}} \in \Delta^{K-1}$, define the marginal types of set B at posterior $\tilde{\boldsymbol{\rho}}$ as $\underline{B}(\tilde{\boldsymbol{\rho}}) \doteq \arg\min_{\theta_i \in B} \langle \tilde{\boldsymbol{\rho}}, \boldsymbol{v_{i\cdot}} \rangle$, and the marginal valuation of set B at posterior $\tilde{\boldsymbol{\rho}}$ as $\underline{v}_B(\tilde{\boldsymbol{\rho}}) \doteq \min_{\theta_i \in B} \langle \tilde{\boldsymbol{\rho}}, \boldsymbol{v_{i\cdot}} \rangle$.

The indirect profit function and the groups of types that buy the good at some optimal price can be defined as follows:

$$\Pi(\tilde{\boldsymbol{\rho}}; \boldsymbol{\alpha}) \doteq \max_{B \subseteq \Theta} \alpha(B) \underline{v}_B(\tilde{\boldsymbol{\rho}}); \qquad \Theta^*(\tilde{\boldsymbol{\rho}}; \boldsymbol{\alpha}) \doteq \arg\max_{B \subseteq \Theta} \alpha(B) \underline{v}_B(\tilde{\boldsymbol{\rho}}).$$

1.4.1 Optimality of full disclosure

The following proposition generalizes the findings of Proposition 1.1. With many types, full disclosure is still optimal if the types are uniformly ordered. In other cases, the suboptimality of full disclosure can be guaranteed for at least some type distributions.

Proposition 1.5 (Optimality of full disclosure)

1. If types are ranked uniformly across the states, full disclosure is optimal for the seller for every type distribution.

2. If types are not ranked uniformly, there is a nonempty, open set of type distributions for which full disclosure is not optimal.

Proof. See Appendix A.3. ■

If the underlying uncertainty cannot change the ranking of the types, then the seller prefers to fully disclose the state so that she can condition the price and the group of types to which she sells on the revealed information. However, if the revealed information can have heterogeneous effects on the different types, withholding some information might bring the expected valuations of certain types closer together such that less surplus might have to be left at these types. The proof of the proposition shows that if the types that are not ranked uniformly across states receive sufficiently high probabilities in the underlying type distribution $\boldsymbol{\alpha}$, then the latter effect is strong enough to guarantee the suboptimality of full disclosure.

The following examples provide further insights concerning the relationship between the type distribution and the optimality of full disclosure. Example 1.4 completely describes the dependence in a simple, three-type, two-state environment. Example 1.5 shows that full disclosure can still be optimal for some type distribution even if no pair of types is ranked uniformly across states.

Example 1.4 Consider an environment with two states and three types with their valuations as given in Table 1.9.

	ψ_1	ψ_2
θ_1	5	3
θ_2	3	5
θ_3	2.5	.5

Table 1.9: Valuations in Example 1.4

Type 3 always has the lowest valuation, and types 1 and 2 are ranked differently in the two states. For $\tilde{\rho}_1 \in [0, 1/2]$, type 2 has the highest expected valuation, followed by types 1 and 3; and for $\tilde{\rho}_1 \in [1/2, 1]$, the order is types 1, 2, and 3. Since the order of the types is constant on these two subintervals, every posterior can be replaced by a convex combination of the endpoints of the subinterval to which it belongs without the seller losing profit. This means that the optimal signal can be constructed using posteriors only from the set $\{0, 1/2, 1\}$.

Therefore, for any interior prior probability $\rho_1 \in (0,1)$, full disclosure is the unique optimal strategy if and only if $\Pi((1/2,1/2);\boldsymbol{\alpha}) < \frac{1}{2}\Pi((1,0);\boldsymbol{\alpha}) + \frac{1}{2}\Pi((0,1);\boldsymbol{\alpha})$. Partial disclosure (or no disclosure for $\rho_1 = 1/2$) is the unique optimal strategy if the reverse inequality is true.

Figure 1.2 illustrates the optimal disclosure policy as a function of the underlying type distribution. The probabilities of the first two types, α_1 and α_2, are measured on the two axes. If type 3 has a sufficiently low probability, full disclosure is not optimal. In this case, the first two types dominate the environment, and the seller has an incentive to hide some information to take advantage of the heterogeneity between these types. In cases in which type 3 has higher probability, the incentives for full disclosure dominate since type 3's valuations are separated from those of type 1 and 2. Finally, if facing a buyer of any of the first two types has a very low probability, then excluding type 3 is not an optimal strategy any more. Consequently, the price is equal to type 3's expected valuation, which is linear in the probability; therefore, all disclosure strategies give the same expected payoff for every interior prior. $\square$

Example 1.5 Consider the four-type three-state environment with valuations as given in Table 1.10.

	ψ_1	ψ_2	ψ_3
θ_1	5	0.5	0.5
θ_2	6	6	0.2
θ_3	0.5	7	7
θ_4	5.5	0.2	8

Table 1.10: Valuations in Example 1.5

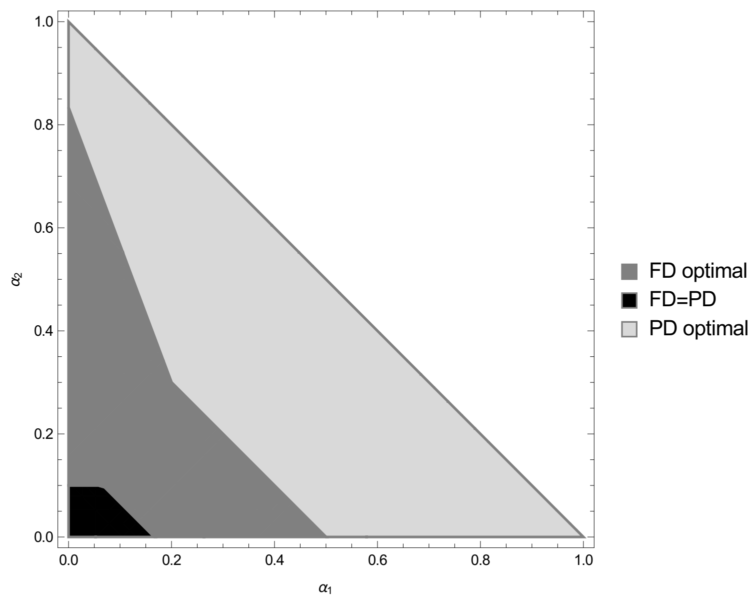


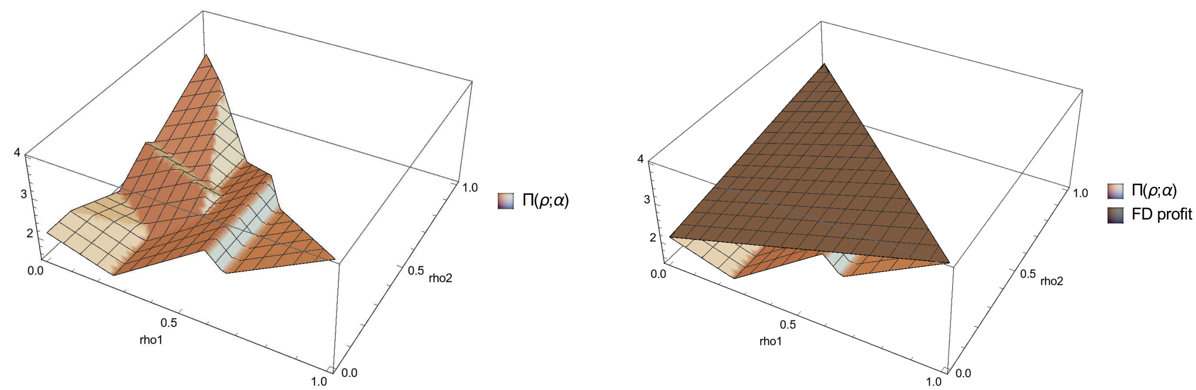
Figure 1.2: Optimal disclosure policy as a function of type distribution

It is easy to verify that preferences are sufficiently mixed and no two types are ranked in the same way for every state. However, for the type distribution $\boldsymbol{\alpha} = (.4, .3, .2, .1)$, as Figure 1.3 shows, full disclosure is an optimal strategy regardless of the prior beliefs. $\square$

1.4.2 Structure of the optimal signal

Let $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(I))$ denote a permutation of the buyer's types. We can think of such a permutation as a weak ordering of the types according to their expected valuations, where the first element corresponds to the type with the highest willingness to pay. Denote the set of all such permutations by Σ.

Let $H(\sigma) \doteq \left\{\tilde{\boldsymbol{\rho}} \in \Delta^{K-1} \colon \langle \tilde{\boldsymbol{\rho}}, \boldsymbol{v}_{\boldsymbol{\sigma(i)}\cdot} \rangle \geqq \langle \tilde{\boldsymbol{\rho}}, \boldsymbol{v}_{\boldsymbol{\sigma(i+1)}\cdot} \rangle, \text{ for every } i = 1, \ldots, I-1\right\}$. This is the set of posteriors at which the weak ranking of the expected valuations of the types coincides with σ. For every $\sigma \in \Sigma$, the set $H(\sigma) \subseteq \Delta^{K-1}$ defines a polytope. These polytopes cover the entire simplex, and their interiors (corresponding to strict rankings) are



(a) Indirect Profit

(b) Optimality of full disclosure

Figure 1.3: Indirect profit $\Pi(\tilde{\boldsymbol{\rho}};\boldsymbol{\alpha})$ and expected profit from full disclosure as a function of the beliefs

mutually disjoint. The following proposition describes the optimal distribution of posteriors based on these polytopes and provides necessary conditions that the groups of types that purchase the product at different posteriors must fulfill in the optimal distribution.

Proposition 1.6

1. The optimal posterior distribution can be constructed using only the extreme points of the polyhedra $(H(\sigma))_{\sigma\in\Sigma}$. These extreme points either are fully revealing or equalize the expected valuations of at least two types.

2. Let $(\boldsymbol{\rho}^{(1)},\ldots,\boldsymbol{\rho}^{(P)};\boldsymbol{\lambda}) \in \Delta^{K-1}\times\ldots\times\Delta^{K-1}\times\Delta^{P-1}$ be an optimal distribution of posterior beliefs. Then for any $\boldsymbol{\rho}^{(p)}$ and $\boldsymbol{\rho}^{(p')}$, where $p,p'\in\{1,\ldots,P\}$, and for any $B\in\Theta^*(\boldsymbol{\rho}^{(p)};\boldsymbol{\alpha})$ and $B'\in\Theta^*(\boldsymbol{\rho}^{(p')};\boldsymbol{\alpha})$, the following hold:

   (i) $B\cap B'\neq\emptyset$;

   (ii) If $\alpha(B)\geqq\alpha(B')$, then $\underline{B}(\boldsymbol{\rho}^{(p)})\nsubseteq B'\setminus\underline{B'}(\boldsymbol{\rho}^{(p')})$;

   (iii) If $\alpha(B)>\alpha(B')$, then $\underline{B}(\boldsymbol{\rho}^{(p)})\nsubseteq B'$.

Proof.

Part 1. Consider a posterior $\tilde{\boldsymbol{\rho}} \in H(\sigma)$ for some $\sigma \in \Sigma$. Since types are ordered uniformly on the convex polytope $H(\sigma)$, the posterior $\tilde{\boldsymbol{\rho}}$ can be decomposed into a convex combination of the extreme points of $H(\sigma)$ without the seller losing profit. Therefore, there must be an optimal distribution of posteriors that assigns positive probability only to the extreme points of the polytopes $(H(\sigma))_{\sigma \in \Sigma}$.

Part 2. Define the functions $g, h\colon [0,1] \to \mathbb{R}$ as follows: for each $\gamma \in [0,1]$,

$$g(\gamma) \doteq \underline{v}_B \left(\gamma \boldsymbol{\rho}^{(\boldsymbol{p})} + (1-\gamma)\gamma \boldsymbol{\rho}^{(\boldsymbol{p}')}\right);$$
$$h(\gamma) \doteq \underline{v}_{B'} \left(\gamma \boldsymbol{\rho}^{(\boldsymbol{p})} + (1-\gamma)\gamma \boldsymbol{\rho}^{(\boldsymbol{p}')}\right).$$

The functions g and h are the lower envelopes of finitely many strictly positive linear functions. Therefore, they are continuous, strictly positive, and weakly concave. To prove the three statements of the second part of the proposition, we will start with indirect assumptions. Then we will construct a posterior belief $\boldsymbol{\rho}^*$ that is a convex combination of $\boldsymbol{\rho}^{(\boldsymbol{p})}$ and $\boldsymbol{\rho}^{(\boldsymbol{p}')}$ such that charging the optimal price at $\boldsymbol{\rho}^*$ leads to a profit strictly higher than the expected profit when $\boldsymbol{\rho}^{(\boldsymbol{p})}$ and $\boldsymbol{\rho}^{(\boldsymbol{p}')}$ are revealed and the optimal prices are charged at each of these posteriors.

Proof of statement 2(i); illustrated in Figure 1.4. Assume to the contrary of the claim that $B \cap B' = \emptyset$. Then the following inequalities must be true:

$$0 < g(0) = \underline{v}_B(\boldsymbol{\rho}^{(\boldsymbol{p}')}) < h(0) = \underline{v}_{B'}(\boldsymbol{\rho}^{(\boldsymbol{p}')});$$
$$g(1) = \underline{v}_B(\boldsymbol{\rho}^{(\boldsymbol{p})}) > h(1) = \underline{v}_{B'}(\boldsymbol{\rho}^{(\boldsymbol{p})}) > 0.$$

Since g and h are both continuous, there exists a $\gamma^* \in (0,1)$ such that $g(\gamma^*) = h(\gamma^*)$. Therefore, at the posterior $\boldsymbol{\rho^*} \doteq \gamma^* \boldsymbol{\rho^{(p)}} + (1-\gamma^*)\boldsymbol{\rho^{(p')}}$, at least all types in B and B' are willing to buy the good if price $g(\gamma^*)$ is charged. It follows that:

$$\begin{aligned}
\Pi(\boldsymbol{\rho^*};\boldsymbol{\alpha}) &\geqq (\alpha(B)+\alpha(B'))g(\gamma^*) = \alpha(B)g(\gamma^*) + \alpha(B')h(\gamma^*) \\
&\geqq \alpha(B)\left(\gamma^* g(1) + (1-\gamma^*)g(0)\right) + \alpha(B')\left(\gamma^* h(1) + (1-\gamma^*)h(0)\right) \\
&> \gamma^*\alpha(B)g(1) + (1-\gamma^*)\alpha(B')h(0) \\
&= \gamma^*\alpha(B)\underline{v}_B(\boldsymbol{\rho^{(p)}}) + (1-\gamma^*)\alpha(B')\underline{v}_{B'}(\boldsymbol{\rho^{(p')}}) \\
&= \gamma^*\Pi(\boldsymbol{\rho^{(p)}};\boldsymbol{\alpha}) + (1-\gamma^*)\Pi(\boldsymbol{\rho^{(p')}};\boldsymbol{\alpha}).
\end{aligned}$$

The second inequality follows from the weak concavity of the functions g and h. The third inequality is strict since we are dropping strictly positive terms.

Proof of statements 2(ii) and 2(iii); illustrated in Figure 1.5. Assume that $\alpha(B) \geqq \alpha(B')$ and that $\underline{B}(\boldsymbol{\rho^{(p)}}) \subseteq B'$. Since expected valuations are linear in the probabilities and there are gaps between the valuations of the marginal types and those of adjacent higher/lower types, there exists $\gamma^* \in (0,1)$ close enough to 1 such that for $\boldsymbol{\rho^*} \doteq \gamma^*\boldsymbol{\rho^{(p)}} + (1-\gamma^*)\boldsymbol{\rho^{(p')}}$, it is true that $\underline{B}(\boldsymbol{\rho^*}) \subseteq \underline{B}(\boldsymbol{\rho^{(p)}})$. For this $\boldsymbol{\rho^*}$, the following holds:

$$\begin{aligned}
\Pi(\boldsymbol{\rho^*};\boldsymbol{\alpha}) &\geqq \alpha(B)\underline{v}_{\underline{B}(\boldsymbol{\rho^{(p)}})}(\boldsymbol{\rho^*}) = \alpha(B)\left(\gamma^*\underline{v}_{\underline{B}(\boldsymbol{\rho^{(p)}})}(\boldsymbol{\rho^{(p)}}) + (1-\gamma^*)\underline{v}_{\underline{B}(\boldsymbol{\rho^{(p)}})}(\boldsymbol{\rho^{(p')}})\right) \\
&\geqq \gamma^*\alpha(B)\underline{v}_B(\boldsymbol{\rho^{(p)}}) + (1-\gamma^*)\alpha(B')\underline{v}_{B'}(\boldsymbol{\rho^{(p')}}) \\
&= \gamma^*\Pi(\boldsymbol{\rho^{(p)}};\boldsymbol{\alpha}) + (1-\gamma^*)\Pi(\boldsymbol{\rho^{(p')}};\boldsymbol{\alpha}).
\end{aligned}$$

The second inequality follows from the two assumptions made at the beginning of the previous paragraph. Moreover, this inequality holds as a strict inequality if $\alpha(B) > \alpha(B')$ or $\underline{B}(\boldsymbol{\rho^{(p)}}) \subseteq B' \setminus \underline{B'}(\boldsymbol{\rho^{(p')}})$. ■

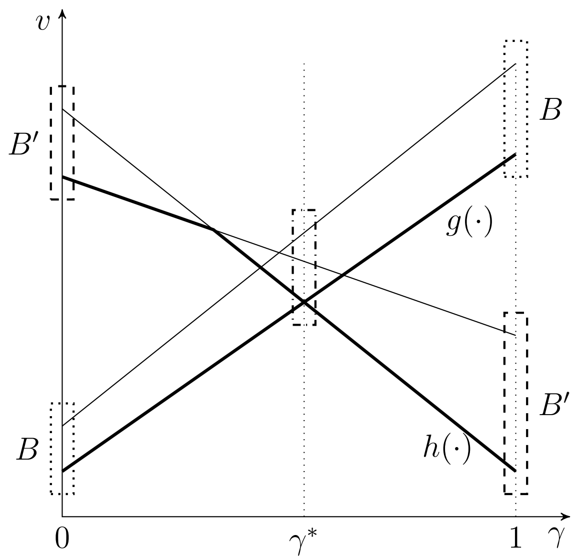


Figure 1.4: $B \cap B' = \emptyset$. By not disclosing any information at γ^*, the seller can sell to both groups with probability 1 such that the expected revenue from each group is strictly higher than the expected revenue when $\gamma = 0$ or $\gamma = 1$ is fully revealed.

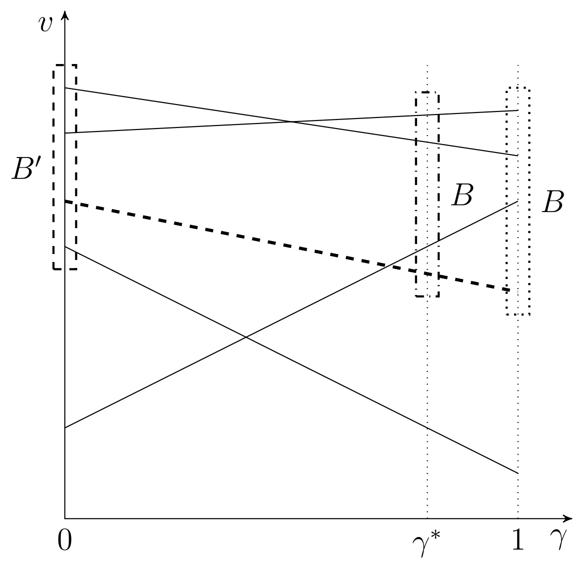


Figure 1.5: $\alpha(B) \geqq \alpha(B')$, marginal type at $\gamma = 1$ (dashed line) buys good also at $\gamma = 0$ but has a higher than marginal valuation. Compared to fully disclosing whether $\gamma = 0$ or $\gamma = 1$, at γ^*, by not revealing any information, the seller can increase the probability of purchase (group B buys the good with probability 1) and also the average price charged (expected valuation of the marginal type at γ^* is above the expectation of the valuation of the marginal type in B' at $\gamma = 0$ and the marginal type in B at $\gamma = 1$).

Proposition 1.6 generalizes two results from the two-type environment to many types. First, as the first part of the proposition shows, mixed beliefs can be used to equalize the expected valuations of some types to capture more surplus. Second, the preferences of the types that purchase the product cannot change too much between different posteriors ((2)(i)). Moreover, the marginal types in a group cannot all be non marginal purchasers at posteriors for which the probability of purchase is weakly lower ((2)(ii)) or cannot all purchase at posteriors for which the probability of purchase is strictly lower ((2)(iii)). Part 2 of Proposition 1.6 implies that in the two-type case the seller always prefers not to reverse the ranking between any pair of posteriors. Otherwise, for such a pair, she must either sell only to the top type at each posterior, which violates (2)(i), or sell to nested sets of buyer types, which violates (2)(ii) or (2)(iii).

1.5 Conclusions

This chapter has studied the optimal design of information disclosure policy in an environment with a single seller, a single indivisible good, and a buyer with private information. In this model, the seller designs a public disclosure policy while keeping in mind its impact on the later screening environment. Moreover, we assume that information might have a horizontal dimension that might affect the ranking of the types by their willingness to pay. These features make the seller's information design problem a nontrivial one: the seller's optimal disclosure policy is not necessarily a binary decision between full and no disclosure, because partially revealing the state can be preferable in many cases to either of those options.

We have shown that the key driving force behind the results is the effect of information on the ranking of the types. More specifically, the seller always prefers to disclose the vertical dimension of information that cannot affect the ranking of the types so she can condition the group of types to which she sells and the price she charges on the revealed information. On the other hand, she might have an incentive to disclose the horizontal dimension only partially to take advantage of its heterogeneous effect on the valuations of different types.

We described the seller's optimal disclosure policy in detail for environments with two types and then offered generalizations of these results to many-type environments. These findings can prove useful for information design problems in real-life environments that exhibit similar characteristics, e.g., setting up review systems in online stores.

There are many possible directions for future research. In the buyer-seller framework, we could relax the unit-demand assumption and let types have different demand functions for the object or we could consider environments with multiple goods and study the joint optimal disclosure problem of the seller. We could also include state-dependent marginal costs to analyze the additional tradeoffs this might create in the information design problem. More importantly, other disclosure problems occur in which initial disclosure is followed by different types of strategic interactions (e.g., the university teaching evaluation and the acquisition examples mentioned in the introduction). It would be interesting to know what the optimal disclosure policy looks like in these environments and whether there are particular features that generalize from the buyer-seller framework. These questions point to some valuable subjects for future research that will certainly enhance our understanding of this relevant and intriguing topic.

Chapter 2

Efficient Resolution of Partnership Disputes

2.1 Introduction

Disputes between partners are often resolved without resorting to dissolution. Depending on the nature of the dispute, a deadlock between partners may lead to the partnership's inevitable dissolution, or it may be overcome through the services of a third party if the latter acknowledges the potential inefficiency associated with dissolution. Indeed, over the past few decades, Alternative Dispute Resolution (ADR) – a variety of dispute resolution methods independent of traditional litigation – has become an integral part of the legal system in many countries. Such methods are perceived as having an advantage over litigation in preserving business relationships, and have become increasingly popular due to concerns over the extensive costs and long delays disputing parties often face taking the litigation route.[1] The American Arbitration Association (AAA), for example – a not-for-profit organization providing ADR services – assures disputing parties that its services "enhance the likelihood of continuing the business relationship".[2]

[1] See, for example, Allison (1990), Lipsky and Seeber (1998), and Shontz et al. (2011).

[2] Similarly, according to the International Centre for Dispute Resolution (ICDR), which administers international arbitration proceedings, "...arbitration assists in minimizing the impact of disputes

Importantly, even if once initiated, arbitration does not explicitly allow for the possibility of avoiding dissolution, its design and associated fees may nevertheless account for the potential inefficiency of dissolution by influencing which disputes are discouraged from entering arbitration, and which are encouraged to do so.

The key novelty in the current work relative to this extensive literature is that dissolution need not be efficient. The resulting endogeneity of dissolution may be a consequence of (i) the arbitrator's capacity to resolve a dispute without dissolution (as in the case of ADR), or (ii) it may be a result of partners' decision whether to enter arbitration in the first place, taking dissolution as given conditional on entry. Regardless of the reason for the endogeneity of dissolution, the goal of this chapter is to shed light on the new trade-offs that arise when it is accounted for, and the implications for the design of arbitration in such environments.

The endogeneity of dissolution opens the door to new questions. Which types of disputes can be resolved efficiently? Is a dispute in a better functioning partnership more or less costly to resolve? What mechanisms can be used to resolve such disputes? Interestingly, the answers to these questions are substantially different when the potential inefficiency of dissolution is taken into account. To study these questions, we consider two-party partnerships with two-sided private information and interdependent valuations.[3] Ownership is dispersed, and the value of the partnership is a function of both agents' independent private information, as well as the severity of the dispute between them (which reduces the partnership's effectiveness). When a partnership is jointly owned, it is natural to assume that one partner's private information also determines the other's valuation from joint ownership. For example, consider a partnership in which each of the two partners is responsible for a separate part of

by resolving them earlier and/or with less disruption to business." See https://www.adr.org/ and https://www.icdr.org/icdr/icdrservices/icdrdisputeresolution.

[3]In the context of joint ventures, approximately 80% of all US joint ventures announced between January 1985 and 2000 in the Joint Ventures and Strategic Alliances database of Thomson Financial Securities Data are two-parent joint ventures (see Hauswald and Hege (2009)).

the business.[4] As Fieseler et al. (2003) point out, in such a scenario, an estimate of the value of the entire business can only be made using information from both parties. Importantly, the value of the partnership also depends on existing frictions, difficulties or costs associated with cooperation or coordination. For instance, consider the following example from Harvard Business Review (November 1986):

> "Two partners owned a company that assembled and marketed an electronic product. One managed the design, marketing, and sales activities. The other handled the procurement, assembly, and finances. At first they disagreed about their customers' needs and their product's features. 'You have overdesigned the product! We can't compete in that market!' one said. 'The market demands an upgraded product! The trouble lies with your failure to assemble to specs!' the other replied... The partners' pressure on each other mounted when the company began to lose money. Antagonism and mutual recrimination sapped energy that might have helped to resolve their market-position and quality concerns."

The key ingredient is that, depending on the partners' private information, the value of the partnership may or may not exceed the partners' valuation for sole ownership of the asset. Therefore, efficient resolution of a dispute requires eliciting the partners' private information in order to determine whether the relationship should be continued or if dissolution is warranted, and if so, who should buy out his associate and at what price.

The analysis first provides a complete characterization of the disputes that can be resolved efficiently. Using this characterization, we show that unless a dispute is sufficiently severe, it cannot be resolved efficiently without incurring a deficit. This impossibility also provides a possible rationale for creating barriers to entry into arbitration.[5] More generally, we show

[4] See Levin and Tadelis (2005), which studies partnerships as revenue-sharing agreements.

[5] In practice, such barriers typically take the form of initial filing fees.

that the severity of the dispute between the partners has a non-monotonic effect on the cost of its efficient resolution. For any initial share allocation, if a dispute is not sufficiently severe, the cost of efficient resolution is necessarily increasing in the severity of the dispute, while for sufficiently large disputes the cost is decreasing. We then show that a first-order stochastic improvement in agents' distribution of valuations – another measure of the partnership's ability to function – may generate lower costs of efficient resolution, unless the probability with which the partnership is dissolved is exactly identical under the two distributions.

Therefore, these results imply that given the endogeneity of dissolution, a dispute in a better functioning partnership may surprisingly be *less* costly to resolve efficiently. To understand the role of endogenous dissolution in this result, observe that if dissolution were taken as given, an improved partnership would simply tighten the partners' participation constraints, making better functioning partnerships more costly to dissolve efficiently. Endogenous dissolution, however, means that an improved partnership, besides this direct effect on the participation constraints, also alters the region of values in which dissolution is efficient in the first place, and consequently the agents' incentive constraints as well. The cost of efficiently resolving a dispute hinges on the interaction between (i) the expected subsidy that must be provided by the arbitrator and (ii) the net expected utility of the partners' "worst-off types" from participation, under an appropriately specified incentive compatible mechanism. The relationship between these different components is at the center of the analysis and crucially depends on the properties of the dispute: its severity, the distribution of the partners' valuations, and the ownership structure.

The analysis then turns to explicitly study endogenous entry into arbitration and indirect mechanisms for efficient resolution of disputes. Assume that having entered arbitration, dissolution is taken as given (i.e., the arbitrator's goal is the standard one considered in the literature: to allocate the asset to the partner who values its sole ownership more).

We augment such arbitration by a preceding stage in which the partners choose whether to enter arbitration or not. Hence, the endogeneity of dissolution stems from the partners' decisions rather than the arbitrator's. We consider a simple class of two-stage games in which the partners first simultaneously choose whether or not to enter arbitration. If neither partner chooses to do so, the partnership remains intact (naturally, in such a case, there are no transfers). If one of the partners chooses to enter arbitration, entry fees are collected (potentially contingent on the entry decisions), and a version of a second-price auction is used to dissolve the partnership. The analysis shows that if, and only if, a dispute can be resolved efficiently, this can be done using such two-stage auctions.[6]

Having studied which disputes can be resolved efficiently, the analysis then considers disputes for which efficient resolution is impossible (i.e., implies a budget deficit). In particular, we focus on the following questions. With the profit from arbitration in mind, should its design create barriers discouraging partners from entering arbitration, or should partners be encouraged to do so? Furthermore, how does the answer to this question depend on the properties of a dispute? To address these questions, we consider a class of direct mechanisms in which the partners' thresholds of dissolution – below which dissolution is triggered – need not be the efficient ones, and derive a condition on disputes necessary and sufficient for the arbitrator's budget to benefit from a locally more/less conservative dissolution threshold for each given partner. Similar to the intuition discussed above for the case of efficient resolution, such direct mechanisms can be shown to be equivalent to two-stage games in which partners first choose whether to initiate arbitration, and if this is the case, efficient dissolution takes place. Importantly, in contrast to the case of efficient resolution, the fees in this case need not be the ones inducing efficient entry.

[6]More precisely, if and only if a dispute can be resolved efficiently, an auction within this class can be constructed which admits a Bayes Nash equilibrium in which the asset is allocated efficiently without incurring a deficit.

For small disputes, which are also the ones for which efficient resolution involves a deficit, more conservative arbitration – discouraging types close to the efficient threshold from initiating arbitration – benefits the arbitrator's budget. In addition, the greater a partner's share, the more the arbitrator's budget benefits from discouraging the partner from entry. In fact, if ownership is sufficiently dispersed, the arbitrator's budget necessarily benefits from discouraging the partner with the greater share from initiating arbitration. Finally, we illustrate these results for uniformly distributed valuations and show that in cases in which efficient resolution yields a deficit, the arbitrator's deficit may be turned into a surplus via a small appropriate change in the dissolution thresholds.

The chapter is organized as follows. Below, we wrap up the introduction with a brief review of the most pertinent literature. Section 2.2 describes the model. Section 2.3 characterizes properties of disputes that can be resolved efficiently. Section 2.4 studies endogenous entry into arbitration and a simple indirect mechanism for efficient resolution. Finally, Section 2.5 extends the results to general partnership functions, and Section 2.6 concludes. Appendix B contains all proofs omitted from the main text.

2.1.1 Related literature

In a seminal paper, Cramton et al. (1987) (henceforth CGK) establish that when ownership is sufficiently close to being symmetric, ex-post efficient dissolution of a partnership (i.e., allocating an asset to the agent with the highest valuation) is possible, but sufficiently asymmetric ownership precludes the possibility of efficient dissolution. In particular, this result contrasts with the impossibility result in Myerson and Satterthwaite (1983), and shows that, with private values, asymmetric ownership rather than asymmetric information is the key factor hindering efficiency.[7] In the context of a public-goods problem with private values, Neeman (1999) shows that efficiency can be obtained only for intermediate property-rights allocations.

Several papers, such as Fieseler et al. (2003), Kittsteiner (2003), Jehiel and Pauzner (2006), Ornelas and Turner (2007), Turner (2013) and Loertscher and Wasser (2016), have studied partnership dissolution in environments with interdependent values. Among other results, these papers identify situations where efficient dissolution may be impossible for any share allocation (see Moldovanu (2002) for an excellent survey). Segal and Whinston (2011) study a general model of Bayesian incentive compatible mechanisms (with either private values or interdependent values) satisfying a certain "congruence" property and describe an ex-ante share allocation for which the interim participation constraints are satisfied. As discussed above, the key departure from the literature in our environment is that the decision to dissolve the partnership is itself endogenous: For some realization of the agents' valuations, efficiency does not call for dissolution.[8]

Salant and Siegel (2016) study efficient allocation of a *divisible* asset between two agents when reallocation is costly. When the asset is initially allocated, the agents' valuations are uncertain. Uncertainty is then resolved and the good may be reallocated at a cost. The surplus from efficient reallocation depends crucially on the curvature of reallocation costs, which determines the optimal initial division of the asset as well as the size of the budget required for the implementation of efficient reallocation. Since reallocation is costly, ex-post efficiency is a function of the initial division of the asset. In contrast, our environment does not assume direct costs of reallocation of shares; rather, such costs are only a result of the need

[7] Schweizer (2006) shows that even if agents' types are not identically distributed, there always exists an initial distribution of shares that permits ex-post efficient dissolution of the partnership.

[8] One exception is Ensthaler et al. (2014) who consider the implications of endogeneity on the implementability of the efficient allocation rule through $k + 1$-price auctions. However, among other differences, in their setting the value of continuing the relationship does not depend on the private information of the partners.

to incentivize the agents while guaranteeing participation. Although the initial allocation of shares plays an important role in determining the possibility of efficient resolution, the ex-post efficient allocation is independent of the initial share allocation.

A related literature has studied specific, widely used mechanisms for partnership dissolution in different environments. McAfee (1992) compares several simple mechanisms for dissolving equal-share partnerships in an independent private values environment. Versions of CGK's $k+1$-auctions are studied in de Frutos (2000), Kittsteiner (2003) and Wasser (2013) (see also references therein) in different settings. De Frutos and Kittsteiner (2008), Brooks et al. (2010) and Landeo and Spier (2013) study versions of the popular Texas Shootout mechanism.[9] Kittsteiner et al. (2012) and Brown and Velez (2016) experimentally compare different partnership dissolution mechanisms under incomplete and complete information, respectively. In a recent work, Van Essen and Wooders (2016) introduce and study a dynamic auction for efficiently dissolving a partnership. While the majority of our analysis takes a mechanism design approach rather than focusing on a particular mechanism, Section 2.4 introduces a simple two-stage mechanism (a second-price auction with endogenous entry) which is shown to resolve a dispute efficiently whenever it is possible to do so.

2.2 Model

A partnership consists of two risk neutral agents $i = 1, 2$ who jointly own an asset. Agent i owns share $r_i \in (0, 1)$ of the asset, with $r_1 + r_2 = 1$. Each agent has private information $\theta_i \in [0, 1]$, which represents the value she attaches to sole ownership of the asset in the absence of the other agent. For example, if the asset corresponds to a firm in which the partners are responsible for different parts of its operation, an agent's type may reflect her resources,

[9] Consistent with the motivation for the current work, Brooks et al. (2010) argue in their conclusion that "certain deadlock situations might be resolved without an actual dissolution of the partnership."

talent, or managerial capability. Each θ_i is drawn independently from a commonly known continuous distribution F on $[0,1]$, with bounded density f, which is assumed bounded away from zero.

Each agent's type θ_i determines her contribution to the partnership. Therefore, the value of the asset when owned jointly depends on both agents' types, and is denoted by $V(\theta_1, \theta_2)$. A dispute between the partners reduces the partnership's value; for example, a dispute may hinder partners' ability to cooperate or give rise to mis-coordination. To facilitate the exposition and illustrate the key tradeoffs that arise due to the endogeneity of dissolution in the simplest way possible, we assume $V(\theta_1, \theta_2) = \theta_1 + \theta_2 - k$, where $k \in [0,1]$ denotes the severity of the dispute between the partners and is publicly observed. (Hence, partner i's payoff from joint ownership of the asset is equal to $r_i\Big(\theta_1 + \theta_2 - k\Big)$).[10] Restricting attention to such a partnership function (as well as more general additive functions) is not necessary for our results. However, it is both a natural representation of profit-sharing arrangements where partners operate independently while sharing profits, and lends more intuition by allowing to obtain closed form expressions of the "worst-off types" (see the discussion in the next section). Section 2.5 extends the results to general increasing, concave partnership functions V satisfying an appropriate single-crossing condition.

A *partnership dispute* is a tuple (r_1, F, k). Depending on each of the agents' valuations for the asset and the severity of their dispute, (ex-post) efficient allocation of the asset may involve either (a) retaining joint ownership of the asset if $\theta_1, \theta_2 \geqq k$, or (b) dissolving the partnership and allocating the asset to agent i if $\theta_i > \theta_{-i}$, and $\theta_{-i} < k$. Let $D \doteq \{d_1, d_2, 0\}$, where d_i denotes the decision to dissolve the partnership and allocate it to i, and 0 the decision to keep the partnership intact.

[10] The analysis focuses on the implementability of efficient allocation of the asset (be it dissolution or retention of the partnership). Other forms of interventions that directly mitigate disputes between the partners (i.e., reduce k through various means) are outside the scope of the analysis.

We study the design of mechanisms with the goal of implementing the efficient allocation rule without the use of external subsidies. By the revelation principle, it is without loss of generality to restrict attention to direct truthful mechanisms. We therefore consider mechanisms of the form (q,t), where $q:[0,1]^2 \to D$ and $t:[0,1]^2 \to \mathbb{R}^2$ denote an allocation and payment rule, respectively.[11] Given a mechanism (q,t) and reported types (θ_1,θ_2), partner i's ex-post net utility is given by $u_i(\theta_1,\theta_2;q,r_i,k)+t_i(\theta_1,\theta_2)-r_i(\theta_1+\theta_2-k)$, where

$$u_i(\theta_1,\theta_2;q,r_i,k) \doteq \begin{cases} \theta_i & \text{if } q(\theta_1,\theta_2)=d_i, \\ 0 & \text{if } q(\theta_1,\theta_2)=d_{-i}, \\ r_i(\theta_1+\theta_2-k) & \text{if } q(\theta_1,\theta_2)=0. \end{cases}$$

If agent i has a true type θ_i, reports θ_i', and believes the other agent reports truthfully, her interim net expected utility is given by

$$U_i(\theta_i',\theta_i;r_i,k) \doteq \int_0^1 u_i\left(\theta_1,\theta_2;q(\theta_i',\theta_{-i}),r_i,k\right)+t_i(\theta_i',\theta_{-i})-r_i(\theta_i+\theta_{-i}-k)\,\mathrm{d}F(\theta_{-i}).$$

Denote $U_i(\theta_i;r_i,k) \doteq U_i(\theta_i,\theta_i;r_i,k)$.

A mechanism is *(interim) incentive compatible* (IC) if $U_i(\theta_i;r_i,k) \geqq U_i(\theta_i',\theta_i;r_i,k)$ for all $\theta_i,\theta_i' \in [0,1]$, *(interim) individually rational* (IR) if $U_i(\theta_i;r_i,k) \geqq 0$ for all $\theta_i \in [0,1]$, and *(ex-ante) budget balanced* (BB) if the arbitrator does not expect to incur positive subsidy payments to the partners: $\mathbb{E}_{\theta_1,\theta_2}\Big(t_1(\theta_1,\theta_2)+t_2(\theta_1,\theta_2)\Big) \leqq 0$. A mechanism is *(ex-post) efficient*

[11] Implicit in this specification is the assumption that the arbitrator must charge the agents before they observe their payoffs. Such an assumption rules out the possibility of full surplus extraction using payments similar to those in Cremer and McLean (1988); see also Mezzetti (2004).

if the allocation rule is the efficient one,

$$q^*(\theta_1,\theta_2) \doteq \begin{cases} 0 & \text{if } \theta_1 > k,\ \theta_2 > k, \\ d_1 & \text{if } \theta_1 > \theta_2,\ \theta_2 < k, \\ d_2 & \text{if } \theta_1 < k,\ \theta_2 > \theta_1, \end{cases} \tag{2.1}$$

illustrated in Figure 2.1.

Definition 2.1 A partnership dispute (r_1, F, k) can be *resolved efficiently* if there exists a mechanism that is efficient, IC, IR and BB.

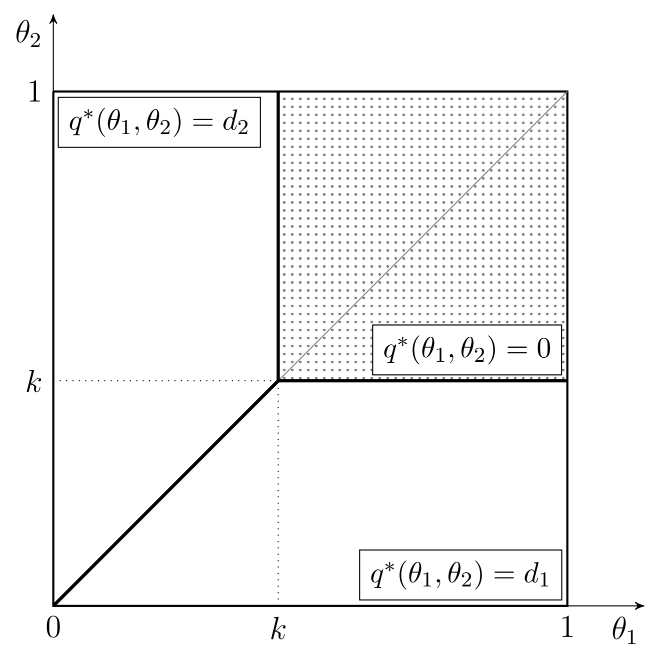


Figure 2.1: Efficient dispute resolution

Some remarks regarding the formulation are in order. First, the setting above implicitly assumes partners cannot simply infer the other's type through observed profits. Indeed, we are concerned with circumstances in which arbitration is necessary in order to resolve such uncertainty. Arbitration may be required when profits are noisy or, as is typically the case in joint ventures, when they will be revealed only at some (unknown) later date. To simplify the exposition, we do not explicitly model these features. Second, in some circumstances the arbitrator might wish to impose finer share-reallocation as a means of providing incentives. The combination of both transfers and arbitrary reallocation of shares does not qualitatively change the results, but greatly complicates the analysis. Furthermore, if the reason for the endogeneity of dissolution is the partners' decision whether to enter arbitration or not, where arbitration consists of efficient dissolution, more general share-reallocation is irrelevant.

2.3 Efficient resolution of disputes

The goal of the analysis in this section is to characterize precisely the class of disputes that can be resolved efficiently, and to illustrate the key new tradeoffs that arise due to the endogeneity of the decision whether to dissolve or retain the partnership. In particular, we study how efficient resolution depends on the effectiveness of a partnership, as measured by the severity of the dispute k and the distribution of agents' private information F. In both cases, the results contrast sharply with the intuition that underlies environments in which dissolution is taken as given.

As a first step, consider the following transfer scheme:

$$t_i^*(\theta_1, \theta_2) \doteq \begin{cases} 0 & \text{if } \theta_i > k,\ \theta_{-i} > k \\ r_i\theta_{-i} & \text{if } \theta_i < k,\ \theta_{-i} > \theta_i \\ -(1-r_i)\theta_{-i} & \text{if } \theta_i > \theta_{-i},\ \theta_{-i} < k \end{cases}, \qquad i = 1, 2. \tag{2.2}$$

Under the mechanism $\Gamma^* \doteq (q^*, t^*)$, if the partnership remains intact, the partners do not make transfers. If the asset is allocated to partner $-i$, partner i receives a transfer that equals his share in the (higher) value generated by $-i$'s sole ownership. Furthermore, partner $-i$ in this case pays an amount equal to i's share in the (lower) value generated by i's sole ownership.

Lemma 2.1 The mechanism $\Gamma^* \doteq (q^*, t^*)$ is efficient, IC and IR.

Proof. See Appendix B.2. ■

The payments t^* guarantee that participation and truthful reporting constitute an ex-post equilibrium. Note that an implication of this result is that unless both partners are either certain the partnership will be dissolved or certain it will not be dissolved, an efficient mechanism that requires no payments be made whenever the partnership remains intact cannot be budget balanced, as dissolution necessarily involves a deficit. In other words, the agents must pay in order to resolve uncertainty about their partnership.[12]

In order to examine the possibility of budget balance, we introduce the following definitions. Given Γ^* and the partnership dispute (r_1, F, k), let

$$\begin{aligned}\mathcal{S}(k) &\doteq \sum_{i=1,2} \int_0^1 \int_0^{\theta_i \wedge k} r_{-i}(\theta_i - \theta_{-i}) \, \mathrm{d}F(\theta_{-i}) \, \mathrm{d}F(\theta_i) \\ &= \mathbb{E}_{\theta_1,\theta_2}\big((\theta_1 - \theta_2)\mathbb{1}(\theta_2 < (\theta_1 \wedge k))\big) \\ &= F(k) \int_k^1 1 - F(\theta) \, \mathrm{d}\theta + \int_0^k F(\theta)\Big(1 - F(\theta)\Big) \, \mathrm{d}\theta \qquad (2.3)\end{aligned}$$

[12] Note that since the values are interdependent, the mechanism Γ^* corresponds to a generalized VCG mechanism (Krishna and Perry (1998)), which in our environment allows for the possibility that neither of the agents are allocated the asset.

denote the *expected subsidy* the arbitrator must incur under Γ^*.[13] Similarly, define the *worst-off type* of each agent i as $\theta_i^*(r_i, k) \in \operatorname{argmin}_{\theta_i \in [0,1]} U_i(\theta_i; r_i, k)$, and let

$$\begin{aligned}\mathcal{L}(r_1, r_2, k) &\doteq U_1(\theta_1^*(r_1, k); r_1, k) + U_2(\theta_2^*(r_2, k); r_2, k) \\ &= k - \int_0^{\theta_1^*(r_1,k)} r_1 - F(\theta_2)\,\mathrm{d}\theta_2 - \int_0^{\theta_2^*(r_2,k)} r_2 - F(\theta_1)\,\mathrm{d}\theta_1 \qquad (2.4)\end{aligned}$$

denote the largest *lump-sum fee* that can be charged from the agents without violating their participation constraints, i.e., the sum of the maximal *participation fees* $U_i(\theta_i^*(r_i, k); r_i, k)$ that the agents are willing to pay. We refer to $\mathcal{L} - \mathcal{S}$ as the *budget surplus* (and to $\mathcal{S} - \mathcal{L}$ as the *budget deficit*).

In order to examine the possibility of BB, it is sufficient to compare the lump-sum fee with the expected subsidy under Γ^*.

Lemma 2.2 Under any efficient, IC and IR mechanism, the worst-off types are equal to $\theta_i^*(r_i, k) = F^{-1}(r_i) \wedge k$. Moreover, the partnership dispute (r_1, F, k) can be resolved efficiently if and only if $\mathcal{L}(r_1, r_2, k) \geqq \mathcal{S}(k)$.

Proof. See Appendix B.3. ■

The proof of Lemma 2.2 follows arguments similar to those in Williams (1999) and Fieseler et al. (2003). Whenever the partnership is dissolved under Γ^* there is necessarily a deficit. However, the participation constraints of the agents typically do not bind, which permits the use of "participation fees" to extract additional surplus from the partners. The highest participation fees that can be charged are those that make the participation constraints of the agents' worst-off types bind. Therefore, a comparison of these highest total

[13]The indicator function $\mathbb{1}$ takes the value 1 if a statement is true and 0 otherwise. See Appendix B.1 for the derivation of the second equality in (2.3) and the equality in (2.4).

participation fees with the expected subsidy determines whether the efficient allocation rule q^* can be implemented without a budget deficit. Finally, a revenue equivalence argument establishes the result. Note that, as in CGK, agents' worst-off types are interior. Intuitively, if a partner's type is high, she not only values the asset more, but it is also more likely that she will receive it. Similarly, the lower a partner's value for the asset the more likely it is that the other partner will be awarded the asset, and the higher the compensation she expects to receive. Hence, the participation constraints of extreme types are more relaxed, resulting in lower information rents for these types.[14]

The following example illustrates the condition above for the simple case of a symmetric partnership and uniformly distributed types.

Example 2.1 Suppose $r_1 = r_2 = \frac{1}{2}$ and $\theta_i \sim U[0,1]$. It can easily be verified that the worst-off types are equal to $\theta_i^*(1/2, k) = k \wedge \frac{1}{2}$, the expected subsidy is equal to $\mathcal{S}(k) = \frac{k}{2} - \frac{k^2}{2} + \frac{k^3}{6}$, and the net expected utility of the worst-off types is

$$\mathcal{L}\left(\frac{1}{2}, \frac{1}{2}, k\right) = \begin{cases} k^2 & \text{if } k < 1/2, \\ k - 1/4 & \text{if } k > 1/2. \end{cases}$$

These functions are plotted in Figure 2.2. Using Lemma 2.2, it can be shown that the partnership dispute can be resolved efficiently if and only if $k \geqq k^* \approx 0.347$ (represented by the third vertical grid line) or $k = 0$. $\square$

Given the necessary and sufficient condition above, we can now study how the possibility or the cost of efficient resolution depends on the effectiveness of a partnership, as measured by the severity of the dispute k and the distribution of the agents' private information F.

[14] A similar intuition underlies the literature on countervailing incentives (see, for example, Lewis and Sappington (1989)).

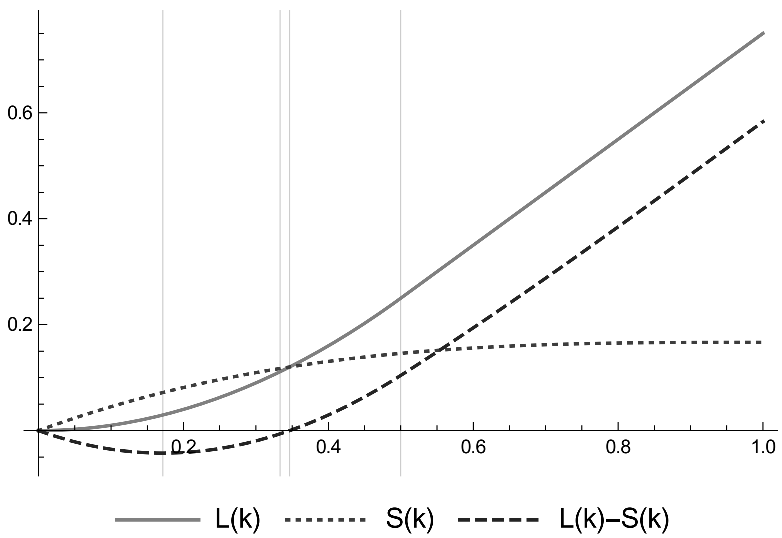


Figure 2.2: Lump-sum fee, expected subsidy, and budget surplus in Example 2.1

2.3.1 Efficient resolution and a partnership's effectiveness

The severity of a dispute k directly impacts a partnership's effectiveness. If the decision to dissolve the partnership were exogenous, an increase in k would simply relax the partners' participation constraints, making it less costly to resolve the dispute. With endogenous dissolution, this observation clearly need not hold, as a change in k additionally alters the region in which dissolution is efficient, and in turn also the worst-off types and their net expected utility.

Before presenting the main results of this section, we illustrate in the corollary below how Lemma 2.2 can be used to derive simple sufficient conditions for (a) the impossibility of efficiently resolving a dispute (independently of the share allocation), and (b) the possibility of efficiently resolving a dispute for equal share partnerships.[15] Note that, trivially, if $k = 0$ then efficient resolution is always possible.

[15]Equal share partnerships constitute the majority of the cases in the samples of two-parent joint ventures in both Veugelers and Kesteloot (1996) and Hauswald and Hege (2009).

Corollary 2.1

1. Any dispute that satisfies $0 < k < \frac{\mathbb{E}\theta + \mathbb{E}(\theta|\theta < k)}{2}$ cannot be resolved efficiently.

2. In a symmetric partnership, i.e., $r_1 = r_2 = \frac{1}{2}$, the dispute can be resolved efficiently if

$$k \geqq \max\left\{ \mathrm{Med}(\theta), F(k)\mathbb{E}\theta + \mathbb{E}\Big(\theta|\theta < \mathrm{Med}(\theta)\Big)\right\}. \tag{2.5}$$

Proof. See Appendix B.4. ■

The condition in part 1 implies that there is an interval $(0, k^*)$ for which the partnership dispute cannot be resolved efficiently. Similarly, if there exists a $k \in [0, 1]$ that satisfies condition (2.5), part 2 identifies a threshold value of k above which disputes in symmetric partnerships can be resolved efficiently.

Example 2.1 (continued) Returning to the setting of Example 2.1, the corollary above establishes that:

1. If $k \in (0, \frac{1}{3})$, the dispute cannot be resolved efficiently, regardless of r (region to the left of the second vertical grid line in Figure 2.2);

2. If $r_1 = r_2 = \frac{1}{2}$ and $k \in [\frac{1}{2}, 1]$, the dispute can be resolved efficiently (region to the right of the fourth vertical grid line in Figure 2.2).

In the case of the symmetric uniform example, the sufficient conditions perform well (the threshold separating the regions where efficient resolution is and is not possible is at the value $k = 0.376$, represented by the third vertical grid line). □

The following proposition shows that, as Corollary 2.1 might suggest, for sufficiently small disputes the deficit is necessarily increasing in k, while for greater disputes the deficit eventually decreases, and the dispute becomes less costly to resolve. An immediate consequence of this result is that since the budget is balanced when there is no dispute, there must exist $\hat{k} \in (0, 1)$ such that any dispute with $k \in (0, \hat{k})$ cannot be resolved efficiently.

Proposition 2.1 For every share allocation (r_1, r_2) and distribution F, the budget surplus $\mathcal{L}(r_1, r_2, k) - \mathcal{S}(k)$ is non-monotonic in k. In particular, there exist $0 < \underline{k} < \overline{k} < 1$ such that the budget surplus is decreasing on $(0, \underline{k})$ and increasing on $(\overline{k}, 1)$.

Proof. See Appendix B.5. ■

The intuition for the effect of a small increase in k is the following. Consider first the case in which the value of k is large. An increase in k raises the expected subsidy as type profiles just above the original threshold start trading and create additional deficit. However, as k becomes large, these additional pairs of types become increasingly similar and therefore have a lower impact on the deficit. Hence, the effect of a small increase in k on the expected subsidy vanishes as k converges to 1. On the other hand, for large enough k values, an increase in k does not change the worst-off types, but increases their net expected utility by worsening the status quo. Since this effect does not vanish as k becomes large, the increase in the lump-sum fee eventually dominates the increase in the expected subsidy.

Consider now the case of low k values. An increase in k results in an increase in the expected subsidy, which does not vanish as k approaches 0. This follows from the fact that the mass of additional trading type pairs is bounded away from 0 (since f is bounded away from 0), and the expected deficit created by these pairs increases as k approaches 0. On the other hand, as opposed to the case of high k, the change in the lump-sum fee for low values of k consists of a change in both the worst-off types and the status quo utilities. If k is small enough, partner i's worst-off type is equal to k. It can easily be checked that, in this case, the net expected utility of i's worst-off type is equal to the (unconditional) expected difference in valuations, $\theta_i - \theta_{-i}$, over the region of θ_{-i} values in which i is awarded the asset. An increase in k increases this difference uniformly for every type of the other partner (below k), therefore the increase in partner i's net expected utility will be proportional to the probability of i being awarded the asset, $F(k)$. Therefore, as k approaches 0, this effect vanishes for both partners, and the change in the lump-sum fee is dominated by the change in the expected subsidy.

Example 2.1 (continued) Consider again the setting of Example 2.1. The derivative of the budget surplus is given by

$$\frac{\partial}{\partial k}\left(\mathcal{L}\left(\frac{1}{2},\frac{1}{2},k\right)-\mathcal{S}(k)\right)=-\frac{1}{2}(k-1)^2+\begin{cases}2k & \text{if } k<1/2,\\ 1 & \text{if } k>1/2.\end{cases}$$

Therefore, the budget deficit increases for $k < 3-2\sqrt{2} \approx 0.172$ and decreases otherwise. This is the region to the left of the first vertical grid line in Figure 2.2. □

We now turn to study the role of another measure of a partnership's effectiveness: the distribution of values F. Observe that the effect of a first-order stochastic improvement creates a shift downwards in the net expected utility curves, resulting in lower participation fees. To see this, first note that a first-order stochastic improvement strengthens the status quo and therefore tightens the participation constraints. Second, whenever a partner is awarded the asset, she pays a higher price more frequently, which also has a negative effect on her net expected utility. Third, in cases in which a partner loses the asset, she gets a higher compensation more frequently, but this positive effect is equivalent to the improvement in the status quo.

On the other hand, an improved type distribution has an ambiguous effect on the expected subsidy. The relationship between the change in the expected subsidy and the change in the lump-sum payment can easily be illustrated for the case in which one of the partners' worst-off types is equal to k and the probability of dissolution does not change. Define first

$$h(\theta_i) \doteq \mathbb{E}_{\theta_{-i}}((\theta_i - \theta_{-i})\mathbb{1}(\theta_{-i} < (\theta_i \wedge k))),$$

which measures the (unconditional) expected difference in valuations over the region in which i is awarded the asset. One can verify that h is nonnegative, and $h'(\theta_i) = F(\theta_i) \wedge F(k)$. Hence, h is strictly increasing, strictly convex on $[0, k]$, and linear on $[k, 1]$. Moreover, $\mathcal{S}(k) = \mathbb{E}_{\theta_i} h(\theta_i)$ and $U(\theta_i^*(r_i, k); r_i, k) = h(k)$ if $\theta_i^*(r_i, k) = k$. The function h is illustrated in Figure 2.3.

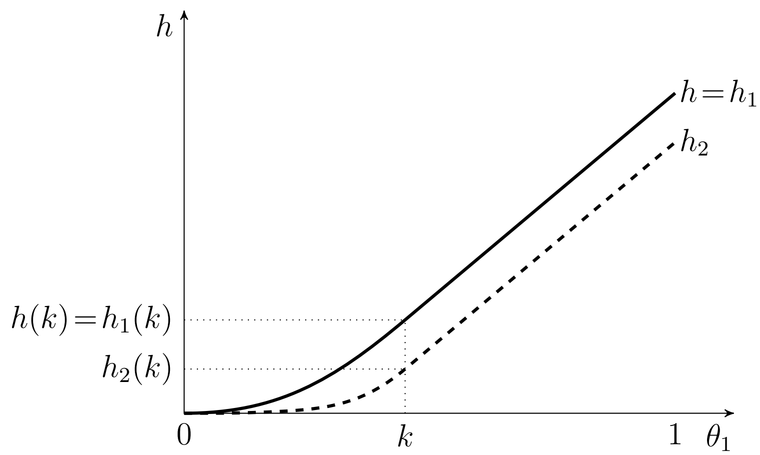


Figure 2.3: Change in h following a first-order stochastic improvement on $[k, 1]$ (h to h_1) and on $[0, k]$ (h to h_2)

Assume that partner i's worst-off type is equal to k. Consider first a first-order stochastic improvement on the interval $[k, 1]$. In this case, h remains unchanged, as the distribution of the θ_{-i} values below k is unaltered. Furthermore, the worst-off type remains the same since the k'th percentile of the distribution is unchanged. Hence, the participation fee (equal to $h(k)$) is unaffected. On the other hand, $\mathcal{S}(k) = \mathbb{E}_{\theta_i} h(\theta_i)$ increases, as the expectation of an increasing function of θ_i is taken with respect to a FOS-improved distribution.

Next, consider a first-order stochastic improvement on the interval $[0, k]$. In this case, h shifts downward, as smaller differences $\theta_i - \theta_{-i}$ are assigned higher mass. Furthermore, the slope of h takes lower values under the FOS-improved distribution for values below k, and remains the same for values above k. This change is represented by the shift from h to h_2 in Figure 2.3. Given the properties of h established above, we can now bound the change in the expected subsidy from above by taking expectations with respect to the improved distribution. In turn, this bound must be smaller than the largest decrease in h over the entire region $[0, 1]$, which occurs at $\theta_i = k$. Since the k'th percentile of the distribution is the same, the worst-off type of partner i remains k, and hence this value is equal to the change in the participation fee of partner i. Since $-i$'s participation fee also decreases following a first-order stochastic improvement, the decrease in the lump-sum fee must dominate the change in the expected subsidy, which makes efficient resolution of the dispute more costly.

The first part of the following proposition shows that this intuition can also be extended to the case in which both worst-off types are below k. In this case, the analysis is complicated by the fact that neither of the participation fees are equal to $h(\theta_i^*(r_i, k))$. The proof is relegated to the Appendix. Interestingly, this intuition need not hold if there is also a decrease in the probability of dissolution, as the second part of the proposition illustrates.

Proposition 2.2 For any two partnership disputes (r_1, F, k) and (r_1, G, k) such that G first-order stochastically dominates F:

1. If the probability of dissolution is the same under F and G, then the dispute in the better functioning partnership (r_1, G, k) is more costly to resolve.

2. Otherwise, the dispute (r_1, G, k) may be less costly to resolve.

PROOF.

Part 1 is relegated to Appendix B.6.

Part 2. To show that the budget surplus can be larger under a first-order stochastically dominating distribution G, consider a sequence of disputes involving symmetric shares and the following sequences of distributions, defined by their pdf's ($n > 5$):

$$f_n(\theta) \doteq \begin{cases} 0.5 & \text{if } \theta \in [0, 0.2 - \frac{1}{n}) \\ 0.5 + 0.1n & \text{if } \theta \in [0.2 - \frac{1}{n}, 0.2) \\ 1 & \text{if } \theta \in [0.2, 1] \end{cases},$$

$$g_n(\theta) \doteq \begin{cases} 0.5 & \text{if } \theta \in [0, 0.2 - \frac{1}{n}) \\ 1 & \text{if } \theta \in [0.2 - \frac{1}{n}, 0.2) \cup [0.2 + \frac{1}{n}, 1] \\ 0.5 + 0.1n & \text{if } \theta \in [0.2, 0.2 + \frac{1}{n}) \end{cases}.$$

The pdf's and cdf's of the two distributions are illustrated in Figures 2.4 (a)-(d).

Assume $k = 0.2$. The median of both distributions is 0.5; therefore, k is the worst-off type of both partners under F_n and G_n. It is easy to see that for every n, the respective cdf's, F_n and G_n, are integrable and G_n first-order stochastically dominates F_n. Using formulas (2.3) and (2.4) for the expected subsidy and the lump-sum payment, it is straightforward to show that the difference in the budget surplus under G_n and F_n is equal to

$$\begin{aligned} \Delta(\mathcal{L} - \mathcal{S}) &= (F_n(0.2) - G_n(0.2))(1 - 0.2) + \int_0^{0.2} F_n(\theta) - G_n(\theta)\,\mathrm{d}\theta \\ &\quad - \left(F_n(0.2)\int_{0.2}^1 F_n(\theta)\,\mathrm{d}\theta - G_n(0.2)\int_{0.2}^1 G_n(\theta)\,\mathrm{d}\theta\right) - \int_0^{0.2} F_n^2(\theta) - G_n^2(\theta)\,\mathrm{d}\theta \\ &\quad - 2\int_0^{0.2} F_n(\theta) - G_n(\theta)\,\mathrm{d}\theta. \end{aligned} \tag{2.6}$$

Since F_n and G_n are identical on $[0, 0.2 - \frac{1}{n}]$ and $[0.2 + \frac{1}{n}, 1]$, they pointwise converge on the set $[0, 0.2) \bigcup (0.2, 1]$ to the same (integrable) function H as $n \to \infty$. Therefore, by the dominated convergence theorem, the change in the budget surplus (2.6) converges to

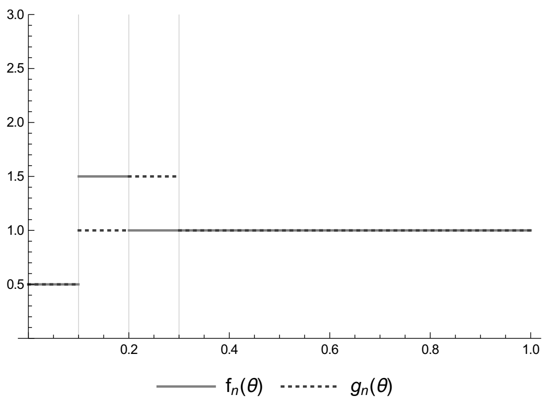
$$\lim_{n\to\infty}(F_n(0.2) - G_n(0.2))\left(1 - 0.2 - \int_{0.2}^{1} H(\theta)\,\mathrm{d}\theta\right) = 0.1 \cdot 0.32 = 0.032 > 0,$$

where H is equal to the cdf of the uniform distribution on $(0.2, 1]$.

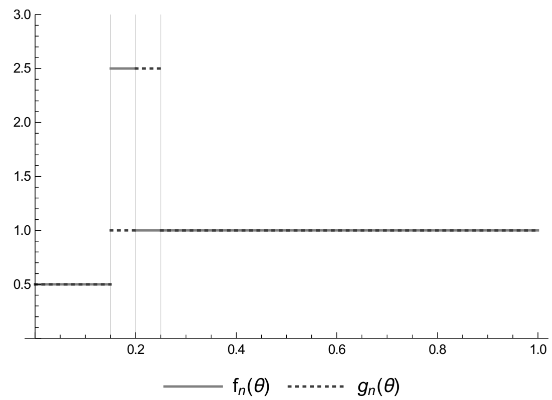
The intuition is the following. The only difference between F_n and G_n is the reallocation of a mass of types from slightly below the threshold to slightly above it. As argued above, the participation fee of a partner whose worst-off type is k is equal to the (unconditional) expected difference in valuations, $\theta_i - \theta_{-i}$, over the region in which i is awarded the asset, at $\theta_i = k$, and the total expected subsidy is the expected value of this unconditional expected difference with respect to θ_i. Since the change is near the threshold, the effect of such a reallocation of mass on the unconditional expected difference in valuations is zero or very small for θ_i types below the threshold, and is continuously increasing for θ_i types above the threshold. Therefore, its effect on the lump-sum fee can be much smaller than the corresponding effect on the expected subsidy. ■

2.4 Endogenous entry into arbitration

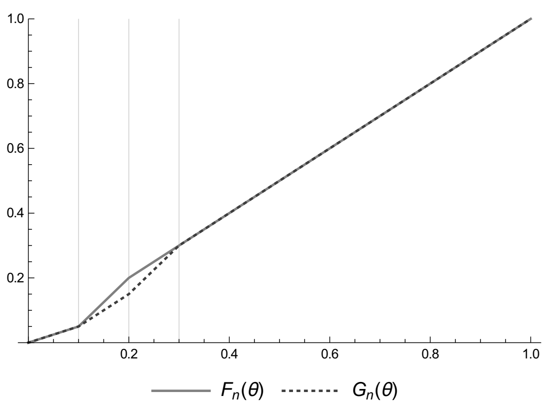
So far, we have considered direct mechanisms for efficient dispute resolution, favoring the interpretation that the endogeneity of dissolution stems from the arbitrator taking into account the dissolution's potential inefficiency. This is indeed consistent, for example, with the goals of ADR. In this section, we explicitly consider an alternative reason for the endogeneity of dissolution – the partners' decision whether or not to enter arbitration in the first place – and study indirect mechanisms for efficient resolution of disputes.



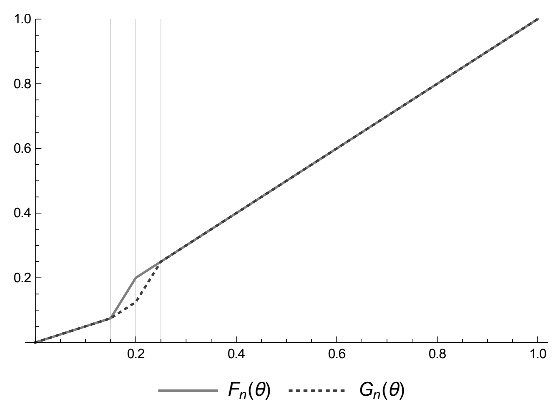
(a) pdf's, $n = 10$



(b) pdf's, $n = 20$



(c) cdf's, $n = 10$



(d) cdf's, $n = 20$

Figure 2.4: The pdf's and the cdf's of the distributions for $n = 10$ and $n = 20$

Suppose that having entered arbitration, dissolution is taken as given. That is, the arbitrator's goal is to allocate the asset to the partner who values its sole ownership more. We augment such arbitration by a preceding stage in which the partners simultaneously choose whether or not to enter arbitration. Hence, the endogeneity of dissolution stems from the partners' decisions, rather than the arbitrator's.

Consider the following class of games.

Definition 2.2 A *second-price auction with endogenous entry* (SAEE) is a two-stage game in which:

1. In stage 1, partners simultaneously choose an action $a_i \in \{D, C\}$, where the action D corresponds to initiating a dispute, and C corresponds to continuing the relationship. If $a_1 = a_2 = C$, no fees are collected, and the game terminates. Otherwise, fees are collected from each agent $i \in \{1, 2\}$, possibly contingent on the agents' announcements, and the game proceeds to stage 2.

2. In stage 2, the partnership is dissolved via a second-price auction. That is, each partner submits a bid b_i, and the partner with the higher bid, $j \in \{1, 2\}$, wins the asset and pays $r_{-j}b_{-j}$. The partner $-j$ who loses the asset receives a compensation of $r_{-j}b_j$. (Note that the difference in the payments between the two cases in which a partner j wins and loses the asset is precisely b_{-j}.)

A *restricted second-price auction with endogenous entry* (R-SAEE) is one in which partners may only submit bids that are consistent with their first-stage action: If a partner chooses $a_i = D$ ($a_i = C$), then she may only submit bids below (above) k.[16]

In the simple game above, partners first decide whether to trigger arbitration. If neither partner chooses to do so, the partnership remains intact (naturally, in such a case, there are no transfers). If one of the partners chooses to enter arbitration, entry fees are collected, and a second-price auction is used to dissolve the partnership.

We now show that the direct mechanism considered in the previous sections and the above defined class of two-stage games – in which dissolution is endogenized through the partners' decision to trigger arbitration – are in fact closely related. Let $\mathcal{E}(T)$ and $\mathcal{E}^R(T)$

[16]For example, inconsistent bids may be replaced with the closest bid consistent with the first-period action, k. Alternatively, a large additional fee may be imposed as a result of such a bid.

denote the set Bayes Nash equilibria (BNE) of the SAEE and R-SAEE, respectively, given the interim expected transfers $T \doteq (T_1(C), T_1(D), T_2(C), T_2(D))$. For off-the-equilibrium path beliefs, we impose the property that each partner believes the other has been following their equilibrium strategies. We say that a SAEE (R-SAEE) resolves a dispute efficiently if there exist fees T such that there is an equilibrium in $\mathcal{E}(T)$ ($\mathcal{E}^R(T)$) that induces the efficient allocation without running a deficit.

Proposition 2.3

1. There exist fees T such that there is an equilibrium in $\mathcal{E}^R(T)$ that induces the efficient allocation. Furthermore, if

$$\int_k^1 \min\{r_1, r_2\}(1-k) - (1-\theta_2)\,\mathrm{d}F(\theta_2) \geqq 0, \tag{2.7}$$

   there exist fees T such that there is an equilibrium in $\mathcal{E}(T)$ that induces the efficient allocation.

2. Consequently, a dispute can be resolved efficiently if and only if there exists a R-SAEE that resolves it, and if (2.7) holds then a dispute can be resolved efficiently if and only if there exists a SAEE that resolves it.

Proof.

Part 1. Consider partner 1's decision problem assuming that partner 2's first period action is always consistent with her type, and her second-period bid is always equal to her true valuation whenever the second-price auction is played.

If arbitration has been triggered, and $\hat{F}_2$ denotes partner 1's belief about partner 2's type (with pdf $\hat{f}_2$), then partner 1's second-period net expected utility from bidding θ_1' when her true type is θ_1 is given by

$$\hat{U}_1(\theta_1', \theta_1; r_1, k, \hat{F}_2) \doteq \int_0^{\theta_1'} \theta_1 - r_2\theta_2 \,\mathrm{d}\hat{F}_2(\theta_2) + \int_{\theta_1'}^1 r_1\theta_2 \,\mathrm{d}\hat{F}_2(\theta_2) - \int_0^1 r_1(\theta_1 + \theta_2 - k) \,\mathrm{d}\hat{F}_2(\theta_2).$$

Computing the first derivative with respect to θ_1', we get

$$\frac{\partial \hat{U}_1(\theta_1', \theta_1; r_1, k, \hat{F}_2)}{\partial \theta_1'} = (\theta_1 - r_2\theta_1')\hat{f}_2(\theta_1') - r_1\theta_1'\hat{f}_2(\theta_1') = (\theta_1 - \theta_1')\hat{f}_2(\theta_1'). \tag{2.8}$$

This expression is nonnegative if $\theta_1' \leqq \theta_1$ and nonpositive if $\theta_1' \geqq \theta_1$. Therefore, in a SAEE, it is a best response for partner 1 to bid her true valuation. Similarly, in a R-SAEE, it is optimal for partner 1 to bid her true valuation if it is consistent with her first-period action, and k (the allowed bid that is closest to her true valuation) otherwise.

Now consider partner 1's decision problem in period 1. If $\theta_1 \leqq k$, partner 1's (re-arranged) first-period incentive compatibility constraint is given by

$$T_1(D) - T_1(C) \geqq F(k)\hat{U}_1(b_1^c, \theta_1; r_1, k, F_{\theta \leqq k}) - \hat{U}_1(\theta_1, \theta_1; r_1, k, F), \tag{2.9}$$

where $F_{\theta \leqq k}$ is the distribution of θ_2 conditional on $\theta_2 \leqq k$, and b_1^c is the second-period best response of partner 1 following a first-period misreport ($b_1^c = \theta_1$ in a SAEE, and $b_1^c = k$ in a R-SAEE). Some algebra on the right-hand side of (2.9) yields

$$\begin{aligned}
T_1(D) - T_1(C) &\geqq F(k)\hat{U}_1(b_1^c, \theta_1; r_1, k, F_{\theta \leqq k}) - \hat{U}_1(\theta_1, \theta_1; r_1, k, F) \\
&= \int_{\theta_1}^{b_1^c} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) + \int_{b_1^c}^{k} r_1\theta_2 \,\mathrm{d}F(\theta_2) - \int_{\theta_1}^{1} r_1\theta_2 \,\mathrm{d}F(\theta_2) \\
&\quad + \int_k^1 r_1(\theta_1 + \theta_2 - k) \,\mathrm{d}F(\theta_2).
\end{aligned}$$

The derivative of the right hand side with respect to θ_1 is equal to $F(b_1^c)-F(\theta_1)+r_1(1-F(k))$. This expression is positive for every $\theta_1 \leqq k$, which shows that constraint (2.9) is the tightest at $\theta_1 = k$:

$$T_1(D) - T_1(C) \geqq F(k)\hat{U}_1(k, k; r_1, k, F_{\theta\leqq k}) - \hat{U}_1(k, k; r_1, k, F) = 0. \tag{2.10}$$

Similarly, if $\theta_1 \geqq k$, the first-period incentive compatibility constraint of partner 1 is

$$T_1(D) - T_1(C) \leqq F(k)\hat{U}_1(\theta_1, \theta_1; r_1, k, F_{\theta\leqq k}) - \hat{U}_1(b_1^d, \theta_1; r_1, k, F), \tag{2.11}$$

where b_1^d denotes the second-period best response of partner 1 following a first-period misreport ($b_1^d = \theta_1$ in a SAEE, and $b_1^d = k$ in a R-SAEE). The right-hand side of (2.11) can be rewritten as follows:

$$\begin{aligned} T_1(D) - T_1(C) &\leqq F(k)\hat{U}_1(\theta_1, \theta_1; r_1, k, F_{\theta\leqq k}) - \hat{U}_1(b_1^d, \theta_1; r_1, k, F) \\ &= -\int_k^{b_1^d} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) - \int_k^{b_1^d} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) - \int_{b_1^d}^1 r_1\theta_2 \,\mathrm{d}F(\theta_2) \\ &\quad + \int_k^1 r_1(\theta_1 + \theta_2 - k) \,\mathrm{d}F(\theta_2). \end{aligned}$$

The derivative of the right-hand side with respect to the first partner's type, θ_1, is equal to $-(F(b_1^d) - F(k)) + r_1(1 - F(k))$. In the R-SAEE case ($b_1^d = k$), this expression is positive, and hence the first-period incentive compatibility constraint is the tightest at $\theta_1 = k$:

$$T_1(D) - T_1(C) \leqq F(k)\hat{U}_1(k, k; r_1, k, F_{\theta\leqq k}) - \hat{U}_1(k, k; r_1, k, F) = 0. \tag{2.12}$$

In the SAEE case ($b_1^d = \theta_1$), the derivative is positive for θ_1 values close to k, and negative for θ_1 values close to 1. Therefore, the first-period incentive compatibility constraint is the tightest either at $\theta_1 = k$ or $\theta_1 = 1$. However, in the latter case, the resulting upper bound for $T_1(D) - T_1(C)$ is negative, which contradicts condition (2.10). To exclude this case, we have to assume that

$$\begin{aligned} 0 &= F(k)\hat{U}_1(k, k; r_1, k, F_{\theta \leqq k}) - \hat{U}_1(k, k; r_1, k, F) \\ &\leqq F(k)\hat{U}_1(1, 1; r_1, k, F_{\theta \leqq k}) - \hat{U}_1(1, 1; r_1, k, F) = \int_k^1 r_1(1-k) - (1-\theta_2)\,\mathrm{d}F(\theta_2). \end{aligned} \tag{2.13}$$

Thus, if $T_1(C) = T_1(D)$ is satisfied, partner 1 has an incentive to trigger a dispute if and only if her type is below k and bid her true valuation in the R-SAEE, as well as in the SAEE whenever (2.13) is satisfied.

Symmetric arguments hold for partner 2, establishing the existence of an equilibrium inducing the efficient allocation for $T_1(C) = T_1(D)$ and $T_2(C) = T_2(D)$ in the R-SAEE, and also in the SAEE if condition (2.7) is satisfied.

Part 2. If $T_1(C) = T_1(D) = T_2(C) = T_2(D) = 0$, then the ex-post payments in the above described equilibrium of the R-SAEE (and of the SAEE if (2.7) is satisfied) are identical to t^* as defined in equation (2.2). Therefore, we can set the first-period interim expected payments to be equal to the maximal participation fees that can be charged in the original one-shot mechanism. The above described equilibirum would then lead to the same budget surplus (or deficit), and to the same interim net expected utilities as in the one-shot case. This, together with revenue equivalence proves the claim. ■

Note that the left hand side of (2.7) is smaller the more mass F has slightly above k. To understand the intuition for condition (2.7) and the distinction between SAEE and R-SAEE, it is useful to consider the incentives of a partner, say partner 1, with type $\theta_1 = 1$. Partner 1 may have an incentive to choose D in the first stage and bid 1 in the second. Such a deviation, which is ruled out under R-SAEE, can be profitable in a SAEE only if partner 2 chooses C in the first stage. Furthermore, conditional on partner 2's bid indeed being above k, this deviation is more profitable the lower 2's bid is, as this implies a lower price to be paid by partner 1, who wins the asset with probability one.

2.4.1 Deficit vs. efficiency

The previous sections have focused on identifying conditions under which disputes can be resolved efficiently, and mechanisms that permit to do so. We now turn to study circumstances in which a dispute cannot be resolved efficiently. In such circumstances, the arbitrator faces a tradeoff between sacrificing efficiency and improving the deficit from arbitration. With the budget from arbitration in mind, we consider the question of whether its design should be conservative, creating barriers discouraging partners from entering (e.g., through high fees or delays), or permissive, encouraging partners to take this step?

In order to study these questions, we consider direct revelation mechanisms with thresholds of dissolution that may potentially differ from k. As in the previous subsection, such mechanisms can be viewed as replicating two-stage games in which dissolution is preceded by a stage in which partners decide whether to enter arbitration. Importantly, the difference here is that the thresholds for dissolution need not be the efficient ones, and are instead chosen by the arbitrator.

Figures 2.5 (a)-(c) illustrate the new allocation and payment rules given an increase in the threshold of partner 1 from k to l_1, the formulas for which can be found in Appendix B.7.[17] It is straightforward to verify that these transfer rules implement this new allocation rule in ex-post equilibrium.

[17] These are slightly different for the $l_1 \geqq k$ and $l_1 \leqq k$ cases. See the Appendix for a discussion regarding the latter case.

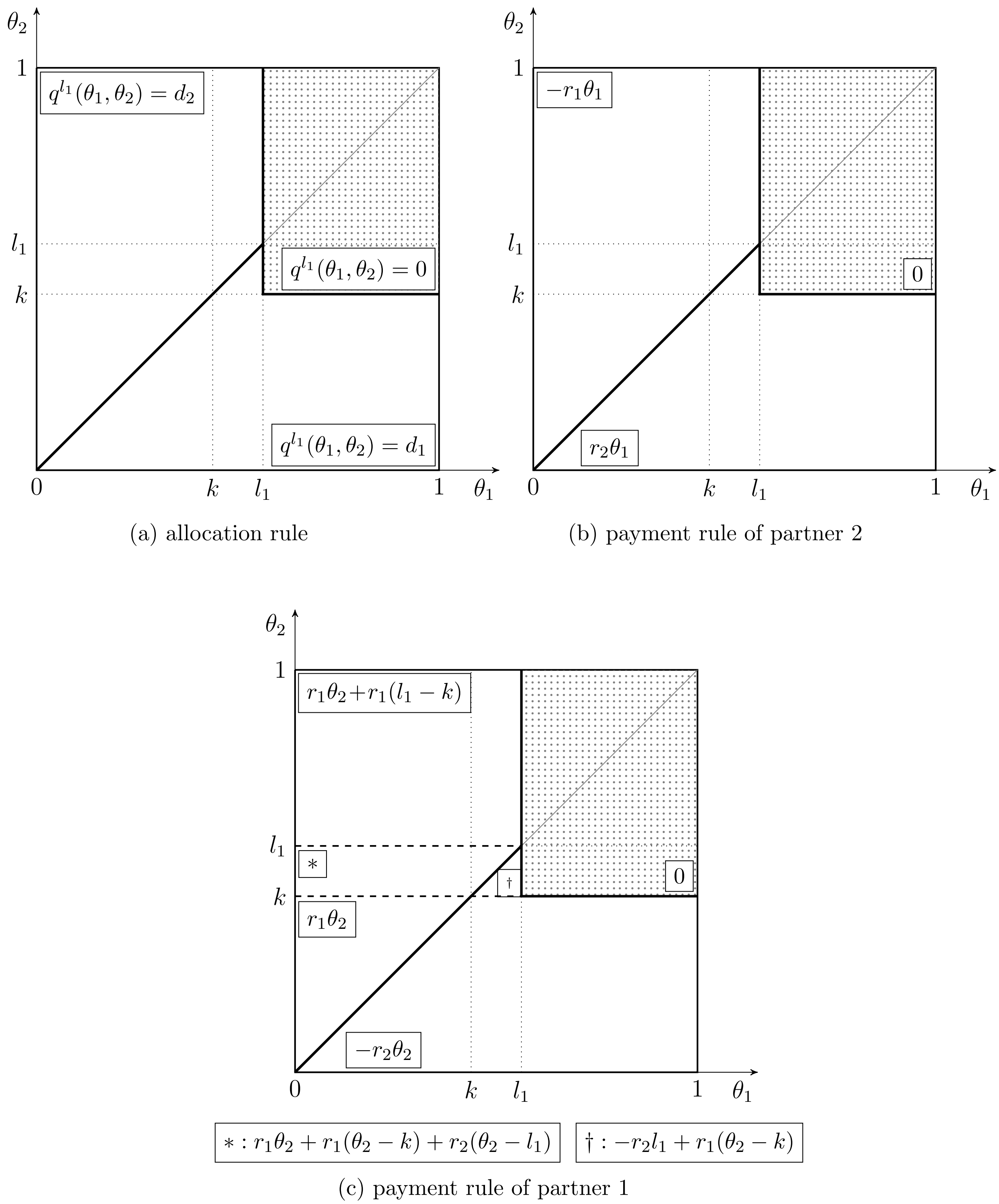


Figure 2.5: Allocation rule and payment rules when partner 1's threshold is raised to $l_1 > k$

The following result shows how the arbitrator can decrease her deficit (or increase her surplus) by varying the threshold of dissolution locally.

Proposition 2.4 A local increase (decrease) in partner i's threshold starting from k decreases the budget deficit if and only if

$$r_i\Big(F(k)(1-F(k)) + f(k)\mathbb{E}((\theta - k) \vee 0)\Big) \leqq (\geqq) \begin{cases} r_i(1-F(k)) & \text{if } r_i \leqq F(k), \\ F(k)(1-r_i) & \text{if } r_i > F(k). \end{cases} \tag{2.14}$$

If $k \neq F^{-1}(r_1)$, $F^{-1}(r_2)$, the effects of simultaneous small changes in the two thresholds can be evaluated independently using (2.14) for both partners.

Proof. See Appendix B.7. ■

Condition (2.14) compares the first-order effects of a slight increase in the threshold (at k) on the expected subsidy (left-hand side) and the participation fee (right-hand side). First, a slight increase in partner i's threshold raises the compensation paid to partner i when she forgoes the asset conditional on partner $-i$'s type being above the new threshold. This higher compensation occurs with probability close to $F(k)(1-F(k))$, therefore it is captured by the first term inside the parenthesis on the left-hand side. Second, raising the threshold increases the probability of dissolution, and therefore it induces a greater trade deficit in the neighborhood of the original threshold of partner i. The increase in the trade deficit is proportional to the mass of partner i types around the original threshold, i.e., $f(k)$, and is represented by the second term inside the parenthesis on the left-hand side. Third, an increase in partner i's threshold also increases her participation fee (the change in partner $-i$'s participation fee is of second order). This change is given by the expression on the right-hand side. In the first case, the original worst-off type of partner i is lower than k.

In the region below k, the only first order consequence of an increased threshold is the higher compensation for selling the asset when $-i$'s type is larger than the new threshold. This affects every type of partner i below k equally, and therefore her worst-off type does not change. The higher compensation (which occurs with probability close to $1-F(k)$), however, increases the maximal fee i is willing to pay. In the second case, the original net expected utility curve of partner i is strictly decreasing up to k (with a negative left derivative at k). Therefore, if the threshold slightly increases, the worst-off type also increases slightly from k. This, together with the downward-sloping net expected utility curve, leads to a weaker positive effect on the partner i's maximal fee than in the first case.

Proposition 2.4 has two immediate implications. First, for small disputes, which as established above are also the ones for which there is a deficit, being more conservative (i.e., discouraging types close to the efficient threshold to initiate a dispute) improves the budget. Second, the higher a partner's share, the more the arbitrator's budget benefits from discouraging her from initiating a dispute. These statements are formalized in the following proposition.

Proposition 2.5

1. There exists a $\tilde{k} > 0$ such that discouraging both partners from initiating arbitration improves the budget surplus for all $k \in (0, \tilde{k})$.

2. If lowering the threshold used for partner i with share r_i improves the budget, then the same holds for any share $r_i' \geqq r_i$. Consequently, the arbitrator's budget benefits more from discouraging the partner with the larger share from initiating arbitration.

3. There exists a $\tilde{r} > 0$ such that if $r_i \in (\tilde{r}, 1)$, then discouraging partner i from initiating arbitration necessarily improves the budget.

Proof.

1. Taking the limit $k \to 0$, the right-hand side of (2.14) converges to 0 in a continuous manner. Moreover, the left-hand side is bounded away from 0 for small k since f is bounded away from 0 and $\mathbb{E}((\theta - k) \vee 0)$ increases as $k \to 0$. Together, these observations guarantee the existence of such a $\tilde{k} > 0$.

2. Dividing both sides of condition (2.14) by $r_i > 0$ we have that decreasing i's threshold improves the budget if and only if

$$F(k)(1 - F(k)) + f(k)\mathbb{E}((\theta - k) \vee 0) \geqq \begin{cases} 1 - F(k) & \text{if } r_i \leqq F(k), \\ F(k)(\frac{1}{r_i} - 1) & \text{if } r_i > F(k). \end{cases} \tag{2.15}$$

   The left-hand side of (2.15) does not depend on r_i while the function (of k) on the right-hand side is weakly lower for larger shares r_i: Assuming $r_i' \geqq r_i$, if $k \geqq F^{-1}(r_i')$, the right-hand side for both shares is the same, $1 - F(k)$; similarly, if $k \in [F^{-1}(r_i), F^{-1}(r_i'))$ then $1 - F(k) \geqq F(k)(\frac{1}{r_i'} - 1)$ is true since $r_i' > F(k)$; finally, if $k < F^{-1}(r_i)$, then $F(k)(\frac{1}{r_i} - 1) \geqq F(k)(\frac{1}{r_i'} - 1)$ holds since $r_i' \geqq r_i$. Thus, condition (2.15) is always satisfied for share r_i' whenever it is satisfied for r_i.

3. If r_i is sufficiently large, the right-hand side of (2.15) approaches 0 as $r_i \to 1$, while the left-hand side remains positive and is independent of r_i. This guarantees the existence of such $\tilde{r} > 0$. ■

Example 2.1 (continued) Returning to the setting of Example 2.1 with equal shares, i.e., $r_1 = r_2 = 1/2$, condition (2.14) implies that reducing efficiency by slightly lowering the threshold of dissolution increases the budget surplus if and only if $k \leqq \sqrt{2} - 1$.

For this symmetric uniform setting, it is straightforward to compute the change in budget surplus for arbitrary (not necessarily local) changes in the thresholds. Assume that the thresholds are changed equally from k to some $l \in [0, 1]$ for both partners. Figure 2.6 illustrates the change in budget surplus for every possible value of l for $k = 0.32$. The figure shows that the positive effect of discouraging types close to the original threshold from initiating dissolution may be strong enough to turn the original budget deficit into a budget surplus. Specifically, decreasing both thresholds to approximately $l = 0.22$ completely balances the budget. This involves making the inefficient decision of retaining the partnership with a probability of $(1-l)^2-(1-k)^2 \approx 14.6\%$. The forgone social surplus is smaller, however, as inefficient decisions occur when they have the lowest impact (close to the threshold). $\square$

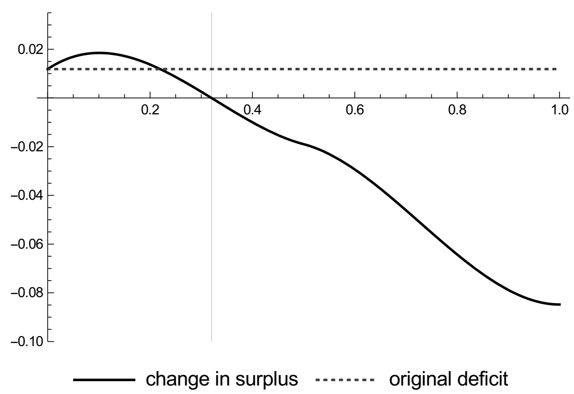


Figure 2.6: Change in surplus and the original deficit in Example 2.1 for $k = 0.32$.

2.5 General partnership functions

The key tradeoffs underlying the analysis above extend to more general environments. Let the value of the partnership $V: [0, 1]^2 \to \mathbb{R}_+$ be any twice differentiable, strictly concave function with $\frac{\partial V(\theta_1,\theta_2)}{\partial \theta_i} \in (0,1)$, $i = 1, 2$.[18] The assumptions guarantee that the value of the partnership is strictly increasing in both partners' types, but the marginal value is strictly decreasing. Furthermore, the difference between a partner's private value θ_i and the value of the partnership $V(\theta_1, \theta_2)$ is strictly increasing in θ_i for all θ_{-i}, which ensures that if it is efficient to assign the asset to partner i then this is also the case for higher types θ_i, holding θ_{-i} fixed. The types are drawn from independent, continuous distributions F_1 and F_2, with positive densities f_1 and f_2.

In Appendix B.8, we show that results similar to those in Sections 2.2-2.3 hold for general partnerships disputes (r_1, F_1, F_2, V). While the generality of the environment no longer permits closed form expressions for the worst-off types of the agents, it is nevertheless possible to derive general properties of these worst-off types and their net expected utility, which permit a characterization of how the different properties of a dispute shape its efficient resolution.

Given the properties of V, there exist threshold types $\underline{\theta}_i(\theta_{-i})$ and $\overline{\theta}_i(\theta_{-i})$ such that: (a) when $\theta_i < \underline{\theta}_i(\theta_{-i})$, or equivalently $\theta_{-i} > \overline{\theta}_{-i}(\theta_i)$, social surplus is maximized when the partnership is dissolved and the asset is allocated to partner $-i$; (b) when $\theta_i \in \left(\underline{\theta}_i(\theta_{-i}), \overline{\theta}_i(\theta_{-i})\right)$, it is efficient to keep the partnership intact. Formally, for any $\theta_{-i} \in [0,1]$, the thresholds are defined by the following equations:

$$\underline{\theta}_i(\theta_{-i}) \doteq \min\left(\{\theta_i \in [0, \theta_{-i}] \colon V(\theta_1, \theta_2) \geqq \theta_{-i}\} \cup \{\theta_{-i}\}\right),$$
$$\overline{\theta}_i(\theta_{-i}) \doteq \min\left(\{\theta_i \in [\theta_{-i}, 1] \colon V(\theta_1, \theta_2) \leqq \theta_i\} \cup \{1\}\right).$$

[18]The assumptions on V are stronger than those required to obtain the results, but greatly simplify the exposition. The results can be generalized to weakly concave functions V with $\frac{\partial V(\theta_1,\theta_2)}{\partial \theta_i} \in [0,1]$. The differentiability of V can also be relaxed.

Given the thresholds $\underline{\theta}_i(\theta_{-i})$ and $\overline{\theta}_i(\theta_{-i})$, the efficient allocation rule, illustrated in Figure 2.7, is given by

$$q^*(\theta_1, \theta_2) \doteq \begin{cases} d_{-i} & \text{if } \theta_i < \underline{\theta}_i(\theta_{-i}) \\ 0 & \text{if } \theta_i \in \left(\underline{\theta}_i(\theta_{-i}), \overline{\theta}_i(\theta_{-i})\right) \\ d_i & \text{if } \theta_i > \overline{\theta}_i(\theta_{-i}) \end{cases}. \tag{2.16}$$

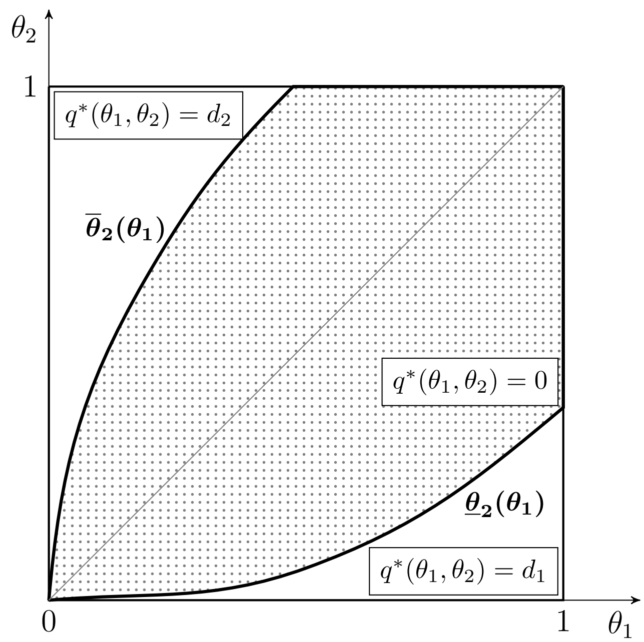


Figure 2.7: Efficient dispute resolution: general V

Denote $\Gamma^* \doteq (q^*, t^*)$, where the transfer rule t^* is defined as

$$t_i^*(\theta_1, \theta_2) \doteq \begin{cases} r_i\theta_{-i} & \text{if } \theta_{-i} > \overline{\theta}_{-i}(\theta_i) \\ 0 & \text{if } \theta_{-i} \in (\underline{\theta}_{-i}(\theta_i), \overline{\theta}_{-i}(\theta_i)) \\ -r_{-i}\overline{\theta}_i(\theta_{-i}) & \text{if } \theta_{-i} < \underline{\theta}_{-i}(\theta_i) \end{cases}. \tag{2.17}$$

Under Γ^*, if the partnership remains intact then the partners do not make transfers. If the asset is allocated to partner $-i$, then partner i receives a transfer that equals his share in the (higher) value generated by $-i$'s sole ownership. Partner $-i$ makes a payment equal to the utility that agent i could have derived from the transferred asset in the best possible effective partnership with $-i$, or in the solely owned asset, whichever is higher.

As in Section 2.3, the feasibility of efficient resolution hinges on the relationship between the net expected utility of the worst-off types and the expected subsidy under Γ^*. We say that the dispute in partnership W is less severe than the dispute in partnership V if $W(\theta_1, \theta_2) \geqq V(\theta_1, \theta_2)$ for all $\theta_1, \theta_2 \in [0, 1]$. In the Appendix (Lemma B.1), we derive properties of the worst-off types of the agents and the net expected utility functions under Γ^*, which can be used to study how a dispute's severity affects its efficient resolution.

A more severe dispute increases both the expected subsidy and the lump-sum fee (Lemma B.3). The relationship between the rate of these two changes determines the overall change in the cost of efficient resolution. Note that the payments, and hence the expected subsidy, depend only on the thresholds of dissolution. The net expected utility, and hence the lump-sum fee, however, depends on other features of the partnership's value (V and W). Therefore, in the particular case in which the region (and consequently the probability) of efficient dissolution is the same for both partnerships, a more severe partnership dispute is less costly to dissolve. If the region of dissolution differs, the relaxation of participation constraints, measured by the change in the lump-sum fee, may be outweighed by the change in the allocation rule, which determines the change in the expected subsidy. Hence, similarly to the intuition discussed in the previous sections, a more severe dispute may be more costly to resolve (Proposition B.1). Finally, we extend the result in Proposition 2.1 by showing that sufficiently small disputes, for which the relative gain from making the efficient allocation decision is small, cannot be resolved without a budget deficit (Proposition B.2).

2.6 Concluding remarks

In this chapter, we have introduced and studied the problem of efficient resolution of partnership disputes in which dissolution need not be efficient. The resulting endogeneity introduces new tradeoffs and yields different predictions relative to environments in which dissolution is taken as given. An implication of this endogeneity is that a dispute in a less effective partnership may be more costly to resolve efficiently. In fact, if a dispute is not sufficiently severe, it cannot be resolved efficiently, and the deficit from efficient resolution is necessarily increasing for small disputes. Endogenizing the decision to enter arbitration in the first place, simple two-stage second-price auctions where partners first choose whether to initiate arbitration achieve efficient resolution if and only if it is feasible. For disputes that cannot be resolved efficiently, whether the arbitrator's budget benefits from locally encouraging or discouraging entry into arbitration hinges on the properties of the dispute.

As a first step in studying the implications of endogenizing the decision to dissolve a partnership, we have focused – in line with much of the partnership dissolution literature – on environments in which the value of a partnership is not a function of the ownership structure. Relaxing this assumption would open the door to different interesting questions. For instance, how might the partners be incentivized to trade shares of an asset with the goal of reaching an optimal ownership structure, given their different valuations for the asset? A related problem is studied in Salant and Siegel (2016) in the context of a divisible asset. Furthermore, a related literature pioneered by Grossman and Hart (1986) and Hart and Moore (1990) studies how the allocation of property rights shapes incentives to invest in the asset's improvement. Studying such incentives in this context seems a particularly interesting direction for future research.

Chapter 3

Stability of Time-Sharing Arrangements in Two-Sided Markets

3.1 Introduction

Fractional matchings are a useful concept for modeling matchings that are probabilistic or have a time dimension. For example, in a marriage market in which agents can be matched to multiple partners over time and have time-invariant preferences, a fractional matching is essentially a table of numbers that specifies the length of time that each pair of agents should spend together. These matchings can be conveniently modeled using linear inequality systems and linear programming.

There are many real-life settings that can be modeled by such fractional matchings. For example, some university programs devote an entire year to internships and practica during which students can take multiple consecutive positions. These programs typically also offer assistance in matching students with potential employers. As another example, we can think of the economics job market interviews at the ASSA meetings, where both job market candidates and potential employers participate in many consecutive interviews.

Despite their usefulness, fractional matchings provide only a reduced-form description of the matching process. Focusing on the time-dimension interpretation from now on, these matchings only describe the length of time that each pair spends together, but not the timing of these matches. The multiplicity of the possible implementing *time schedules* makes it a nontrivial task to extend concepts such as stability to this environment.

Stability requires the non-existence of a blocking pair, i.e., a pair of agents that are both matched to partners that they desire less than each other. However, for a given fractional matching, the existence of a blocking pair can depend on the implementing time schedule. Because of the above mentioned multiplicity, even if there exists a schedule without a blocking pair (i.e., the underlying fractional matching satisfies *weak stability*), there might be other schedules in which a blocking pair can arise for a positive amount of time. Choosing such a weakly stable fractional matching might be unattractive in practical applications for several reasons. First, the mechanism designer might have to exercise more control and be involved more closely in the scheduling process if there is a chance that a blocking pair can arise, instead of letting the market decide about the timing (e.g., using a first-come first-served sign-up method). Second, in many typical real-life settings, previously unexpected events might occur during the matching process. These impose new constraints on the timing of matchings, and the remaining matches might have to be rescheduled in a way that is consistent with both the original underlying fractional matching and the additional constraints.[1] This rescheduling might also lead to a blocking pair if there are unstable continuation schedules.

Roth et al. (1993) define a stronger stability condition that solves this problem. *Strong stability* requires that for every pair, either the man or the woman spend no time at all with someone less desired than the other. Since strong stability makes it impossible for such a

[1]For example, during the 2018 ASSA Meetings in Philadelphia, a significant fraction of the incoming flights was canceled, and many interviews had to be rescheduled.

potentially blocking pair to exist by definition, blocking pairs cannot arise in any of the time schedules that are consistent with a strongly stable matching.

In this chapter, we argue that strong stability can be too demanding since it typically excludes fractional matchings that otherwise can be implemented only in a stable way. More precisely, a potentially blocking pair can be harmless in the sense that no matter what the implementing time schedule is, its members can never be matched to inferior partners *simultaneously*. Another limitation of strong stability is that, as Roth et al. (1993) show in their paper, in a strongly stable matching no agent can spend time with more than two partners. This seems to be a limiting feature especially in larger markets, and excludes many fractional matchings that cannot be implemented in an unstable way.

Based on these observations, we define a new stability concept for fractional matchings, called *C-stability*, that is more general than weak stability but less general than strong stability. The new concept is equivalent to the non-existence of unstable implementing time schedules. We characterize this condition both geometrically and in terms of the underlying preferences.

The geometric characterization reveals the relationship between the set of fractional matchings satisfying our stability concept, the polytope of matchings and the polytope of weakly stable fractional matchings (also called the *stable matching polytope*). We show that a fractional matching satisfies C-stability if and only if the smallest face of the matching polytope that contains it has only stable vertices. Therefore, the set of C-stable matchings is the union of the faces of the matching polytope that are also faces of the stable matching polytope.

Building on these results, we obtain necessary and sufficient conditions on the underlying preferences that must be fulfilled for a candidate fractional matching to satisfy our notion of stability. Using graph theory, we can represent potential pairs in the marriage market by a rectangular array of nodes, in which rows and columns correspond to men and women, respectively. Directed edges connect nodes in every row and every column, representing the preferences of men and women over the other side of the market.

For a fractional matching that lies on an edge or a diagonal of the matching polytope (hence, it can be expressed as a convex combination of two vertices), the pairs that are matched for positive time define a set of disjoint preference cycles and isolated nodes in the graph representing the marriage market. These cycles and nodes partition the sets of men and women, which can be used to derive the following necessary and sufficient conditions for C-stability for all possible pairs of agents in the marriage graph:

(i) Every preference cycle should be directed. *(directed cycle condition)*

(ii) For every pair of a man and a woman whose members belong to different cycles or isolated nodes, the man should rank the woman below his least preferred partner, or the other way around. *(external stability)*

(iii) For every pair of a man and a woman whose members belong to the same cycle, if the man and the woman like each other more than their least preferred partners, they should also like each other less than their most preferred partners. *(internal stability)*

This characterization result can be immediately extended to fractional matchings in which no agent spends time with more than two partners, which also highlights the relationship between strong stability and C-stability. The directed cycle and external stability conditions are equivalent to the requirement defined by strong stability for these groups of pairs. However, the internal stability condition is weaker since it also incorporates the underlying combinatorial restrictions of the problem of finding implementing time schedules.

The characterization obtained for general fractional matchings uses the necessary and sufficient conditions derived for edges and diagonals. It applies them to the sets of preference cycles and isolated nodes that are defined by a subset of the positive coordinates of the matching matrix in a way that partitions the sets of men and women. Contrary to strong stability, agents can have more than two partners over time in C-stable matchings.

3.1.1 Related literature

Although the properties of stable matchings in the marriage market (Gale and Shapley (1962)) show a remarkable similarity to those in the assignment game (Shapley and Shubik (1971)), the two models were originally analyzed using different mathematical tools. While the assignment game was traditionally modeled by linear programming, discrete mathematical tools were applied to analyze marriage markets.

The first paper that modeled the stable marriage problem as a linear programming problem was Vande Vate (1989). The paper defined a system of inequalities whose extreme points contain all the stable matchings. Rothblum (1992) modified Vande Vate's LP formulation to obtain an inequality system such that the stable matchings are *exactly* the extreme points of the polytope defined by the feasible solutions. Roth et al. (1993) used this linear programming formulation to explain the remarkable similarity between the properties of the optimal assignment problem and the stable marriage problem.

A by-product of the linear programming formulation is the existence of fractional solutions that can be interpreted as time-sharing arrangements or lotteries over non-fractional matchings. Given the strong technical tools available to model fractional matchings, it is not surprising that they have been frequently used to analyze a wide range of practical matching problems such as school choice lotteries or two-sided search. For some recent examples, see Kesten and Ünver (2015), Lauermann and Nöldeke (2014), and Echenique et al. (2013).

Roth et al. (1993) emphasized that weak stability of fractional matchings does not guarantee that only stable implementing schedules exist. Their strong stability concept offers a solution to this issue by completely excluding pairs who might have an incentive to increase the time spent together and decrease the time spent with their less preferred partners.

Some works investigated the geometric and combinatorial properties of the stable matching polytope. Balinski and Ratier (1997) used graph theory to describe stable matchings and to obtain a characterization of the faces of the stable matching polytope. The graph representation of markets that we also use in this chapter appeared in their work first. The special case of our directed cycle condition in the case of fractional matchings generated by neighboring vertices is what they call *comparability* of neighboring stable vertices.

Cycles defined by preferences were also used in the book Gusfield and Irving (1989). They constructed an algorithm involving preference cycles that can be used to find all stable matchings in a marriage market.

Given a stable fractional matching, Abeledo et al. (1996) and Teo and Sethuraman (1998) suggested very simple procedures that can be used to obtain stable implementing time schedules in which matchings are ordered in a linear way.

Our work is related to these papers but goes further. Instead of describing geometric properties of the stable matching polytope, it investigates the relationship between the matching polytope and the stable matching polytope and gives necessary and sufficient conditions for the stability of whole faces of the former.

The chapter is organized as follows. Section 3.2 provides the basic definitions and some existing results about weak stability and strong stability of fractional matchings. Section 3.3 formally defines our proposed stability concept, C-stability. Sections 3.4 and 3.5 describe the geometric and preference-based characterizations of C-stability. Section 3.6 concludes the chapter. Appendix C contains the proofs that are omitted from the main text.

3.2 Preliminaries

Geometry

A *polyhedron* is defined as the set of solutions to a linear inequality system. A *polytope* is a bounded polyhedron. Polytopes are convex sets that can be characterized equivalently by their defining inequalities and as the convex hull of their extreme points.

If we replace some inequalities with equality constraints, the result will be a subpolytope, a *face* of the original polytope. A 0-dimensional face is called a *vertex*, a 1-dimensional is called an *edge* of the original polytope. The vertices are exactly the extreme points of the polytope. It is straightforward to verify that the set of all faces of a polytope, ordered by set inclusion, forms a lattice.

Two vertices are called *neighboring* if their convex hull is an edge of the polytope. The line segment connecting two vertices is a *diagonal* of the polytope if the two vertices are not neighboring.

The *relative interior* of a set is its interior within its affine hull. The *relative boundary* is its boundary within its affine hull.

Matching

Our environment is a regular marriage market, consisting of men and women. The finite disjoint sets of men and women are denoted by $M = \{m_1, \ldots, m_{|M|}\}$ and $W = \{w_1, \ldots, w_{|W|}\}$, respectively. Each man $m \in M$ and each woman $w \in W$ has a strict preference ordering defined on the other side of the market, denoted by $\succ_m$ and $\succ_w$, respectively. We assume for simplicity that every pair (m, w) is acceptable to both m and w, for each $m \in M$ and $w \in W$.

Definition 3.1 (Matching polytope $\boldsymbol{P}$) Given a set of men M and women W, the matching polytope P is defined as the set of matrices $\mathbf{x} \in \mathbb{R}^{|M|\times|W|}$ satisfying

$$\begin{aligned} \sum_{j\in W} x_{mj} &\leqq 1 \qquad \text{for all } m \in M, \\ \sum_{i\in M} x_{iw} &\leqq 1 \qquad \text{for all } w \in W, \\ x_{mw} &\geqq 0 \qquad \text{for all } (m,w) \in M \times W. \end{aligned}$$

We call the solutions to this problem *fractional matchings* and the integer solutions *matchings*.[2]

A fractional matching $\mathbf{x}$ can be interpreted either as a time assignment or as a collection of matching probabilities. In the time assignment interpretation, the value x_{mw} specifies the *length of time* m and w must spend together. The total available time for matchings is normalized to 1, and we assume constant preferences over time without discounting. In the matching probability interpretation, a fractional matching is simply a reduced form where x_{mw} specifies the *probability* with which m and w are matched. For the sake of simplicity, from now on, we will use the time assignment interpretation; most of what follows remains true if fractional matchings represent probabilities.

In the definition, the first two groups of inequalities state that no one can be matched for longer than the total available time. The third group of constraints means that assignment times are always non-negative.

A fractional matching is not a complete description of the matching process since it does not define *the timing* of the matchings. A *time schedule implementing a fractional matching*

[2] Similarly to many papers in the literature, we index the elements in x by men and women, which is a slight abuse of notation. However, it greatly simplifies the expressions in the chapter. Since we can assume that row i represents m_i and column j w_j, this should create no confusion.

$\mathbf{x}$ is defined as a convex combination of matchings.[3] Such a convex combination does not fully specify the timing either, it tells us only which pairs are matched at the same time. Since the order of the matchings does not directly influence the agents' preferences, and our focus is on stability (whether a blocking pair can exist for some time), this definition is sufficient for our analysis.

It has been shown in the literature that the extreme points (vertices) of the polytope P are *exactly* the possible matchings between men and women.[4] Since every point in P can be expressed as a convex combination of vertices, *this result guarantees the existence of an implementing time schedule for every fractional matching.*

Definition 3.2 (Stable matching polytope S) Given a set of men M and women W, the stable matching polytope S is defined as the set of matrices $\mathbf{x} \in \mathbb{R}^{|M|\times|W|}$ satisfying

$$\begin{aligned} \sum_{j\in W} x_{mj} &\leqq 1 \quad \text{for all } m \in M, \\ \sum_{i\in M} x_{iw} &\leqq 1 \quad \text{for all } w \in W, \\ x_{mw} &\geqq 0 \quad \text{for all } (m,w) \in M \times W, \\ \sum_{j\succ_m w} x_{mj} + \sum_{i\succ_w m} x_{iw} + x_{mw} &\geqq 1 \quad \text{for all } (m,w) \in M \times W. \end{aligned}$$

We call the solutions to this inequality system *weakly stable fractional matchings*[5] and the integer solutions *stable matchings.*

[3]When describing time schedules, we restrict attention to convex combinations in which all the weights are positive.

[4]See e.g., Roth et al. (1993).

[5]The literature typically refers to such matchings as "stable." In this chapter, we use "weak stability" to make it easier to distinguish this definition from the other two notions of stability we consider, strong stability and C-stability.

The last condition is the generalization of the stability condition of matchings. Reorganizing it gives

$$\sum_{j \succ_m w} x_{mj} + \sum_{i \succ_w m} x_{iw} \geqq 1 - x_{mw} \quad \text{for all } (m, w) \in M \times W.$$

This condition simply means that m and w must spend at least as much time with better partners as the time they do not spend together. For matchings (where the elements of the matrix $\mathbf{x}$ are binary), this is equivalent to the original definition of stability. For fractional matchings, this condition is necessary for the non-existence of a blocking pair (m, w). Since a stable matching always exists in a marriage market, S is never empty.[6]

A time schedule implementing a fractional matching is *stable* if it is a convex combination that involves only stable matchings. The following proposition shows that the weak stability of a fractional matching and the existence of a stable implementing time schedule are equivalent.

Proposition 3.1 (Rothblum (1992)) The extreme points of the stable matching polytope S are exactly the stable matchings.

3.2.1 Strong stability

If we interpret fractional matchings as time allocations, then the original stability definition might not be strong enough. Although there is a stable implementing schedule for every weakly stable fractional matching, other time schedules with a blocking pair might still exist (the last constraint is only necessary for the non-existence of a blocking pair). The following example (taken from Roth et al. (1993)) illustrates this observation.

[6] See e.g., Roth and Sotomayor (1992).

Example 3.1 (Roth et al. (1993)) Consider the following preferences (every preference list goes from the most preferred potential partner to the least preferred one):

Men:				Women:			
$m_1:$	w_1	w_3	w_2	$w_1:$	m_3	m_2	m_1
$m_2:$	w_2	w_1	w_3	$w_2:$	m_1	m_3	m_2
$m_3:$	w_3	w_2	w_1	$w_3:$	m_2	m_1	m_3

Given these preferences, there are three stable matchings:

$$\mathbf{u^1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{u^2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{u^3} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Consider the following two convex combinations:

$$\mathbf{y} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

The fractional matching $\mathbf{y}$ is weakly stable since it can be implemented by a stable time schedule (first convex combination). However, there also exists another time schedule (second convex combination) that implements $\mathbf{y}$ for which a blocking pair exists at every point in time. □

The issue created by this multiplicity is solved by *strong stability*, defined in Roth et al. (1993).

Definition 3.3 (strong stability) A weakly stable fractional matching $\mathbf{x}$ is strongly stable if for all $m \in M$ and $w \in W$ it is true that

$$\left(1 - \sum_{i \succeq_w m} x_{iw}\right)\left(1 - \sum_{j \succeq_m w} x_{mj}\right) = 0.$$

Strong stability requires that for every pair, either the man or the woman spend no time at all with someone less desired than the other one. Therefore, it completely excludes potentially blocking pairs. It is easy to see that strong stability is not satisfied in Example 3.1 (e.g., both m_1 and w_3 spend time with less preferred partners).

A strongly stable matching always exists since a stable matching always exists in the marriage problem and stability trivially implies strong stability for matchings.

Next, we summarize the necessary and sufficient conditions for strong stability.

3.2.2 Necessary and sufficient conditions for strong stability

Roth et al. (1993) derived the following necessary and sufficient conditions for strong stability.

Proposition 3.2 (Necessary conditions for strong stability) If $\mathbf{x}$ is a strongly stable fractional matching, then

(i) for all $w \prec_m w''$, such that $x_{mw}, x_{mw''} > 0$, it is true that if $w \prec_m w' \prec_m w''$ for some $w' \in W$, then $y_{mw'} = 0$ in every weakly stable fractional matching $\mathbf{y}$,

(ii) for all $m \prec_w m''$, such that $x_{mw}, x_{m''w} > 0$, it is true that if $m \prec_w m' \prec_w m''$ for some $m' \in M$, then $y_{m'w} = 0$ in every weakly stable fractional matching $\mathbf{y}$.

Proposition 3.2 immediately implies that in every strongly stable matching, every man and woman can have *at most two different partners*. This might be an unappealing property in larger markets with many agents on both sides.

Proposition 3.3 (Sufficient conditions for strong stability) Let $\mu_M(w)$ and $\mu_W(m)$ denote the worst partners that w and m can get in any stable matching. A weakly stable fractional matching $\mathbf{x}$ is strongly stable if either of the following holds:

(i) for all $m \in M$ and $w, w', w'' \in W$ such that $w \prec_m w' \prec_m w''$ and $x_{mw''} > 0$, it is true that $\mu_M(w') \succ_w m$,

(ii) for all $w \in W$ and $m, m', m'' \in M$ such that $m \prec_w m' \prec_w m''$ and $x_{m''w} > 0$, it is true that $\mu_W(m') \succ_m w$.

3.3 C-stability

Strong stability excludes all potentially blocking pairs by definition. However, in many cases this might be too demanding: there might be pairs of agents who both spend time with someone less desired than the other one, but can never do that *at the same time*, no matter what the implementing time schedule is. Such agents would have to compare their losses from blocking (giving up a better partner for the other blocking agent) and their gains (replacing worse partner with the other blocking agent). Even if such an agreement is mutually desired by the blocking agents, enforcing it might still be an issue since the last agent to make a sacrifice would have an incentive not to return the favor and keep his existing partner. In many real settings, a commitment device for such an agreement might not be available (e.g., writing a contract on blocking might be illegal, highly unethical, or simply too costly), and such agents can never form a blocking pair.[7] These observations motivate the following definition.

Definition 3.4 (C-stability) A fractional matching is C-stable if it cannot be expressed as a convex combination of matchings that involves at least one unstable matching.

As we discussed, convex combinations contain all the information on timing that is important for stability. Therefore, C-stability is equivalent with the non-existence of unstable implementing time schedules. The next proposition shows the relationship between weak stability, C-stability and strong stability.

Proposition 3.4

1. If $\mathbf{x}$ is strongly stable, then it is C-stable.
2. If $\mathbf{x}$ is C-stable, then it is weakly stable.

PROOF.

1. Assume that $\mathbf{x}$ is not C-stable. Then $\mathbf{x}$ can be expressed as a convex combination of matchings that involves at least one unstable matching. Therefore, there is a positive amount of time for which a blocking pair exists. Strong stability is violated.
2. This direction is trivial since every fractional matching can be expressed as a convex combination of matchings. If this cannot involve unstable matchings, then the fractional matching must be weakly stable, as well. ∎

The next example shows that the converse of the two statements in Proposition 3.4 is not true.

[7]The same holds if we interpret fractional matchings as arrays decribing matching probabilities.

Example 3.2 (Counterexamples) Consider again the setting of Example 3.1:

Men:				Women:			
$m_1:$	w_1	w_3	w_2	$w_1:$	m_3	m_2	m_1
$m_2:$	w_2	w_1	w_3	$w_2:$	m_1	m_3	m_2
$m_3:$	w_3	w_2	w_1	$w_3:$	m_2	m_1	m_3

Given these preferences, it is straightforward to show that exactly the following matchings are stable:

$$\mathbf{u^1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{u^2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{u^3} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

1. *(C-stability does not imply strong stability.)* Consider the following stable fractional matching:

$$\mathbf{x} = \frac{1}{2}\mathbf{u^1} + \frac{1}{2}\mathbf{u^2} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

This fractional matching is C-stable: since $x_{13} = x_{21} = x_{32} = 0$, and everyone is matched for the whole time, the same must hold for every matching in a convex combination. However, only the matchings $\mathbf{u^1}$ and $\mathbf{u^2}$ fulfill this requirement. On the other hand, $\mathbf{x}$ is not strongly stable since both members of the pair (m_1, w_3) spend positive time with less desired partners: $\sum_{j \prec_{m_1} w_3} x_{m_1 j} = 1/2$ and $\sum_{i \prec_{w_3} m_1} x_{i w_3} = 1/2$.

2. *(Weak stability does not imply C-stability.)* Consider again the following two convex combinations from Example 3.1:

$$\mathbf{y} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

The first convex combination contains only stable matchings, the second only unstable ones. Hence, the fractional matching $\mathbf{y}$ is weakly stable but not C-stable. □

3.4 Geometric properties of C-stability

The goal of this section is to describe C-stability geometrically. From now on, assume for simplicity that $|M| = |W| = n$. This assumption implies that in every stable matching (consequently, in every weakly or strongly stable fractional matching), everyone must be matched for the whole time.[8] Moreover, if we want to express a weakly stable fractional matching as a convex combination of matchings, we need to use matchings that satisfy the same property. Therefore, we can restrict attention to the following polytopes.

[8]See e.g., Roth and Sotomayor (1992).

Definition 3.5 (Matching polytope P') The matching polytope P' is defined as the set of matrices $\mathbf{x} \in \mathbb{R}^{n \times n}$ satisfying

$$\begin{aligned}
\sum_{j \in W} x_{mj} &= 1 \quad \text{for all } m \in M, \\
\sum_{i \in M} x_{iw} &= 1 \quad \text{for all } w \in W, \\
x_{mw} &\geqq 0 \quad \text{for all } (m,w) \in M \times W.
\end{aligned}$$

Definition 3.6 (Stable matching polytope S') The stable matching polytope S' is defined as the set of matrices $\mathbf{x} \in \mathbb{R}^{n \times n}$ satisfying

$$\begin{aligned}
\sum_{j \in W} x_{mj} &= 1 \quad \text{for all } m \in M, \\
\sum_{i \in M} x_{iw} &= 1 \quad \text{for all } w \in W, \\
x_{mw} &\geqq 0 \quad \text{for all } (m,w) \in M \times W, \\
\sum_{j \succ_m w} x_{mj} + \sum_{i \succ_w m} x_{iw} + x_{mw} &\geqq 1 \quad \text{for all } (m,w) \in M \times W.
\end{aligned}$$

It is still true that the extreme points of these two polytopes are exactly the matchings and the stable matchings, since P' and S' are just the faces of P and S for which the first two groups of inequalities are binding.

Since the set of faces of a polytope, ordered by set inclusion, forms a lattice, for every fractional matching $\mathbf{x}$, we can define *the smallest face of P containing* $\mathbf{x}$. Denote it by $F(\mathbf{x})$. The following lemma shows how we can construct the face $F(\mathbf{x})$ for a given fractional matching $\mathbf{x}$.

Lemma 3.1 (Smallest containing face) Let $\mathbf{x}$ be a fractional matching. Then

$$F(\mathbf{x}) = \text{conv}\left\{\mathbf{u} \in P' : \mathbf{u} \text{ is a vertex of } P', \text{ and } x_{mw} = 0 \Rightarrow u_{mw} = 0, \text{ for every } m, w\right\},$$

where conv denotes the convex hull of a set.

Proof. The set on the right-hand side defines the face of the polytope P' for which those inequality constraints are binding that are also binding for $\mathbf{x}$. Therefore,

$$F(\mathbf{x}) \subseteq \text{conv}\left\{\mathbf{u} \in P' : \mathbf{u} \text{ is a vertex of } P', \text{ and } x_{mw} = 0 \Rightarrow u_{mw} = 0, \text{ for every } m, w\right\}$$

must be true. However, the set on the right-hand side cannot be a strict superset of $F(\mathbf{x})$. Otherwise, we could find an inequality constraint that is not binding for the face on the right-hand side but must be binding for all elements in $F(\mathbf{x})$. In this case, there would exist a pair (m, w) such that $x_{mw} > 0$ and $y_{mw} = 0$ for all $\mathbf{y} \in F(\mathbf{x})$, contradicting $\mathbf{x} \in F(\mathbf{x})$. ■

Lemma 3.1 suggests a simple way to compute the vertices of the smallest containing face for a given fractional matching: fix the coordinates at which the fractional matching takes the value of zero and assign 1 or 0 to the remaining coordinates in a consistent way (such that the row and column sums of the resulting matrix are 1 everywhere).

The next lemma will be useful for establishing our characterization result.

Lemma 3.2 Let $\mathbf{x}$ be a fractional matching and $F(\mathbf{x})$ be the smallest face containing $\mathbf{x}$. Then $\mathbf{x}$ can be expressed as a convex combination of the vertices of $F(\mathbf{x})$ such that all coefficients are positive.

Proof. See Appendix C.1 ■

Using Lemmas 3.1 and 3.2, we can prove the following characterization result:

Theorem 3.1 (Geometric characterization) A fractional matching $\mathbf{x}$ is C-stable if and only if every vertex of $F(\mathbf{x})$ is a stable matching.

PROOF. Assume first that $\mathbf{x}$ is a C-stable fractional matching. From Lemma 3.2 we know that $\mathbf{x}$ can be expressed as a convex combination of all vertices in $F(\mathbf{x})$ such that every weight is positive. The definition of C-stability immediately implies that every vertex of $F(\mathbf{x})$ must be stable.

For the other direction, observe that every convex combination that yields $\mathbf{x}$ can involve vertices only from $F(\mathbf{x})$. To see this, assume that $\mathbf{x}$ can be expressed as a convex combination that involves a vertex outside $F(\mathbf{x})$, say $\mathbf{v}$, with a positive weight. This means that one of the non-negativity constraints that are binding for all elements of $F(\mathbf{x})$ is not binding for $\mathbf{v}$. Therefore, there must be a pair m, w such that $x_{mw} = 0$ but $v_{mw} > 0$. However, $\mathbf{v}$ appears with positive weight in the convex combination that is equal to $\mathbf{x}$. Therefore, $x_{mw} > 0$ should be true, a contradiction. ∎

Theorem 3.1 also implies that if $\mathbf{x}$ is C-stable, then each vector in the face $F(\mathbf{x})$ is C-stable. (For each $\mathbf{y} \in F(\mathbf{x})$, the smallest containing face is either $F(\mathbf{x})$ itself or a subface of it. In both cases, all the vertices of the face $F(\mathbf{y})$ are stable.) Moreover, since S' is generated by some vertices of P', if a face of P' has only stable vertices, it must be a face of S' as well. These two observations lead to the following corollary.

Corollary 3.1 The set of all C-stable fractional matchings is the union of those faces of the matching polytope P' that are also faces of the stable matching polytope S'.

The set of C-stable matchings is not necessarily convex. All stable matchings are C-stable and their convex hull is exactly S'. However, as we have seen, not all weakly stable fractional matchings are C-stable in general.[9]

3.5 Preference-based characterization of C-stability

In this section, we derive conditions on the underlying preferences that are both necessary and sufficient for C-stability. We use a geometric approach. First, we discuss the simple case of a fractional matching lying on an edge of the matching polytope. Then we generalize the result to fractional matchings lying on diagonals and finally to general fractional matchings.

Before we do this, the first thing to check is whether the necessary conditions for strong stability (Proposition 3.2) are still necessary for C-stable matchings. As the following example shows, this is not the case since an agent can spend time with more than two partners in a C-stable matching, which contradicts the implication of Proposition 3.2.

Example 3.3 Consider the preference profile

Men:					Women:				
$m_1:$	w_1	w_2	w_3	w_4	$w_1:$	m_2	m_1	m_3	m_4
$m_2:$	w_2	w_1	w_3	w_4	$w_2:$	m_3	m_1	m_2	m_4
$m_3:$	w_3	w_2	w_1	w_4	$w_3:$	m_1	m_2	m_3	m_4
$m_4:$	w_1	w_2	w_3	w_4	$w_4:$	m_1	m_2	m_3	m_4

and the fractional matching

[9] The same logic shows that the set of strongly stable fractional matchings is not necessarily convex either.

$$\mathbf{x} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By fixing the 0 coordinates ($x_{14} = x_{23} = x_{24} = x_{31} = x_{34} = x_{41} = x_{42} = x_{43} = 0$), and assigning the remaining 0 and 1 values in a consistent way, we can obtain the vertices generating $F(\mathbf{x})$:

$$\mathbf{u^1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{u^2} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{u^3} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is easy to verify that all three vertices are stable. Therefore, the fractional matching $\mathbf{x}$ is C-stable. However, at $\mathbf{x}$, the first man and the second woman spend positive time with three potential partners. Therefore, this fractional matching violates the necessary condition for strong stability. □

We need to make another preliminary observation regarding the relationship between vertices. Given two arbitrary vertices (say $\mathbf{u^1}$ and $\mathbf{u^2}$), the special structure of the matrices (there is exactly one "1" element in every row and column) allows us to transform the matrix of one vertex into that of the other by performing a *row permutation*. Denote this permutation by $\Sigma_{\mathbf{u^1},\mathbf{u^2}}$.[10] It is a basic result in algebra that each permutation can be written as the product

[10] To simplify notation, we will not distinguish between row permutations and functions applying the row permutations to matching matrices. This should create no confusion since the object in the argument of a permutation will uniquely determine the sense in which we are using it.

of mutually disjoint *cyclic permutations* that are uniquely determined.[11] This means that there is a unique set $\{\sigma_1, \dots, \sigma_L\}$ of disjoint cycles such that

$$\mathbf{u^2} = \Sigma_{\mathbf{u^1},\mathbf{u^2}}(\mathbf{u^1}) = (\sigma_1 \cdot \ldots \cdot \sigma_L)(\mathbf{u^1}).$$

We use this result extensively in this secion.

3.5.1 Fractional matchings lying on edges

Let $\mathbf{u^1}$ and $\mathbf{u^2}$ denote the two neighboring vertices defining the edge. In this case, the row permutation $\Sigma_{\mathbf{u^1},\mathbf{u^2}}$ is a cycle itself. To see this, assume to the contrary that there are at least two disjoint cycles (σ_1 and σ_2) in the product. By taking the transformation $\sigma_1(\mathbf{u^1})$ or $\sigma_2(\mathbf{u^1})$, we do not change the binding non-negativity conditions (the coordinates that are zero in *both* $\mathbf{u^1}$ and $\mathbf{u^2}$ remain zero if we apply any of the cycles in the product). This means that $\mathbf{u^1}$, $\sigma_1(\mathbf{u^1})$, and $\sigma_2(\mathbf{u^1})$ must all belong to the smallest face containing both $\mathbf{u^1}$ and $\mathbf{u^2}$. Since the two cycles are disjoint, we have three different vertices lying on the same edge, which is a contradiction.[12]

Let μ_1 and μ_2 be two functions that return the partners of the agents at the matchings $\mathbf{u^1}$ and $\mathbf{u^2}$, respectively.[13] If σ denotes the unique row cycle transforming $\mathbf{u^1}$ into $\mathbf{u^2}$, the previous observation immediately implies the following connection between the two neighboring matchings:

[11] A cyclic permutation is a permutation of a subset of rows that has a single nontrivial cycle. An example for a cyclic permutation is the following transformation: row 1 is replaced by row 2, row 2 is replaced by row 3, and row 3 is replaced by row 1.

[12] Note that in the 2x2 and 3x3 cases, all pairs of vertices are neighboring since we need at least four rows for the simplest product of disjoint cycles.

[13] For $k = 1, 2$, the function $\mu_k \colon M \cup W \to M \cup W$ is such that $\mu_k(m) \doteq \{w \in W : u^k_{mw} = 1\}$ for all $m \in M$ and $\mu_k(w) \doteq \{m \in M : u^k_{mw} = 1\}$ for all $w \in W$.

$$\mu_2(m_k) = \begin{cases} \mu_1(m_k) & \text{, if } k \notin \sigma; \\ \\ \mu_1(m_{\sigma(k)}) & \text{, if } k \in \sigma. \end{cases}$$

Using Theorem 3.1, we can derive the following necessary condition for the C-stability of a fractional matching lying on an edge of the matching polytope.[14]

Proposition 3.5 (comparability) Assume that a fractional matching $\mathbf{x}$ lies on an edge of the matching polytope P' that is generated by $\mathbf{u^1}$ and $\mathbf{u^2}$. If $\mathbf{x}$ is C-stable, then the following must be true:

The men and women who have different partners at $\mathbf{u^1}$ and $\mathbf{u^2}$ must have the opposite preferences over the two matchings. If σ is the cycle transforming $\mathbf{u^1}$ into $\mathbf{u^2}$, then either

- $\mu_1(m_k) \succ_{m_k} \mu_2(m_k)$ and $m_k \prec_{\mu_1(m_k)} m_{\sigma^{-1}(k)}$ for all $k \in \sigma$ (every man prefers his partner at $\mathbf{u^1}$ to the one at $\mathbf{u^2}$ and every woman does it the other way), or
- $\mu_1(m_k) \prec_{m_k} \mu_2(m_k)$ and $m_k \succ_{\mu_1(m_k)} m_{\sigma^{-1}(k)}$ for all $k \in \sigma$ (the other way).

Proof. Denote the cycle transforming $\mathbf{u^1}$ into $\mathbf{u^2}$ by $\sigma = (a_1\, a_2\, \ldots\, a_K)$. From Theorem 3.1, we know that both $\mathbf{u^1}$ and $\mathbf{u^2}$ must be stable for $\mathbf{x}$ to be C-stable.

Take man m_{a_1} and assume without loss of generality that m_{a_1} strictly prefers his partner under $\mathbf{u^1}$ to that under $\mathbf{u^2}$, i.e., $\mu_1(m_{a_1}) \prec_{m_{a_1}} \mu_2(m_{a_1}) = \mu_1(m_{a_2})$. Then

$$\mu_2(\mu_1(m_{a_2})) = m_{a_1} \succ_{\mu_1(m_{a_2})} m_{a_2} = \mu_1(\mu_1(m_{a_2}))$$

must be true, otherwise m_{a_1} and $\mu_1(m_{a_2})$ would form a blocking pair at $\mathbf{u^1}$. By the same reasoning, it is also true that $\mu_1(m_{a_2}) \prec_{m_{a_2}} \mu_2(m_{a_2}) = \mu_1(m_{a_3})$, otherwise m_{a_2} and $\mu_1(m_{a_2})$

[14] Balinski and Ratier (1997) calls this condition comparability of neighboring matchings. An alternative proof can be given using Theorem 10 of their paper.

could block the matching $\mathbf{u^2}$. We can continue applying the same reasoning for every man and woman involved in the cycle σ, which proves the claim. ■

The condition can be illustrated using graph theory. Consider an $n \times n$ array of nodes representing the possible pairs that can form in the marriage market, where rows correspond to men and columns to women. In the row of every man m_i, directed edges connect the nodes, pointing from the less preferred pair to the more preferred pair according to m_i's preferences. Edges in columns are defined analogously.

The original row cycle $\sigma = (a_1, \ldots, a_K)$ defines the following cycle *in the graph*:

$$\begin{aligned}\rho \doteq &((m_{a_1}, \mu_1(m_{a_1})), (m_{a_1}, \mu_2(m_{a_1})), (m_{a_2}, \mu_1(m_{a_2})), \\ &(m_{a_2}, \mu_2(m_{a_2})), \ldots, (m_{a_K}, \mu_1(m_{a_K})), (m_{a_K}, \mu_2(m_{a_K}))).\end{aligned}$$

The cycle ρ connects pairs who spend positive time together in an alternating way: it has two nodes in each row and each column involved in the cycle. Call such a cycle *preference cycle*. Then, the condition of Proposition 3.5 can be rephrased as follows.

Corollary 3.2 (directed cycle condition) Assume $\mathbf{x}$ is a fractional matching lying on an edge generated by the vertices $\mathbf{u^1}$ and $\mathbf{u^2}$. Let σ denote the unique row cycle transforming $\mathbf{u^1}$ into $\mathbf{u^2}$ and let ρ be the preference cycle induced by σ. If $\mathbf{x}$ is C-stable, then the following condition must hold:

(DC) The cycle ρ must be a *directed cycle* of preferences.

The following example illustrates the directed cycle condition.

Example 3.4 Consider again the setting of Example 3.3. We have seen that the following fractional matching is weakly stable by construction:

$$\mathbf{x} \doteq \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The row cycle transforming the first vertex into the second one is given by (132). Since both vertices are stable and adjacent, $\mathbf{x}$ is C-stable.

Figure 3.1 (a) shows the *graph representation* of the preferences. Arrows point from the less preferred to the more preferred partners, crossed and gray nodes represent the two vertices).[15] Figure 3.1 (b) shows the preference cycle defined by the row permutation (132). It is indeed a directed cycle, which is consistent with the statement of Corollary 3.2. □

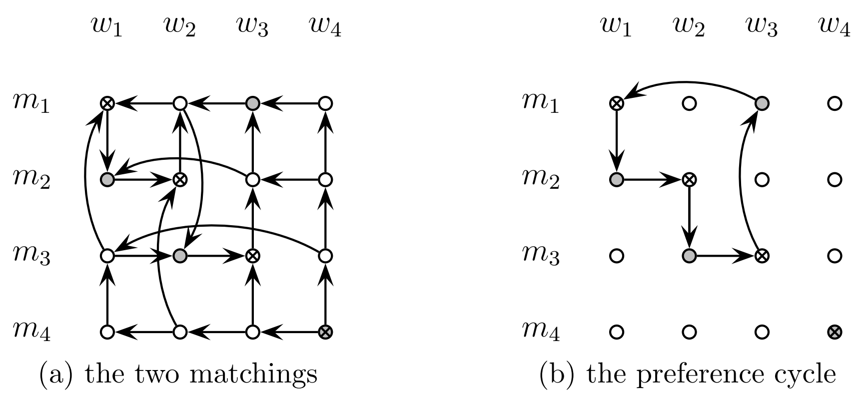


Figure 3.1: Preferences and the directed cycle condition in Example 3.4

The following example shows that the directed cycle condition is only necessary, but not sufficient for C-stability. The additional conditions that are needed for the characterization will be presented in Section 3.5.3.

Example 3.5 (DC is only necessary) Consider again the setting of Example 3.4, but assume that $m_3 \succ_{w_1} m_2$ and $w_1 \succ_{m_3} w_2$ hold. The condition (DC) is still satisfied. However,

[15] Arrows implied by transitivity are omitted.

given the new preferences, the gray matching cannot be stable since man m_3 and woman w_1 would block it. $\square$

3.5.2 Fractional matchings lying on diagonals

Let $\mathbf{x}$ be a fractional matching lying on a line segment connecting two non-neighboring vertices $\mathbf{u^1}$ and $\mathbf{u^2}$. Since these vertices are not neighboring, other vertices must also belong to the smallest face containing $\mathbf{x}$. We generalize (DC) by constructing the smallest containing face $F(\mathbf{x})$ using the cycle factorization $\Sigma_{\mathbf{u^1},\mathbf{u^2}} = \sigma_1 \cdot \ldots \cdot \sigma_L$. In the next lemma, we show that every subset of these cycles applied to $\mathbf{u^1}$ defines a vertex in the face $F(\mathbf{x})$, and every vertex of this face can be constructed by this method.

Lemma 3.3 (smallest face generated by a diagonal) The following must be true for the smallest face containing $\mathbf{x}$:

$$F(\mathbf{x}) = \operatorname{conv}\left\{ \left(\prod_{\lambda \in \Lambda} \sigma_\lambda \right) (\mathbf{u^1}) : \Lambda \subseteq \{1, \ldots, L\} \right\}$$

Proof. Remember that the smallest containing face is defined by the binding non-negativity constraints. Since $\mathbf{x}$ is a convex combination of $\mathbf{u^1}$ and $\mathbf{u^2}$, if a non-negativity constraint is binding for $\mathbf{x}$ (a coordinate is zero), the same must hold for both $\mathbf{u^1}$ and $\mathbf{u^2}$.

The vertices $\mathbf{u^1}$ and $\mathbf{u^2}$ are not neighboring by assumption. Therefore, the factorization of $\Sigma_{\mathbf{u^1},\mathbf{u^2}}$ must involve at least two disjoint cycles. Any cycle σ_l involved in the factorization transforms a set of rows of $\mathbf{u^1}$ into the same set rows of $\mathbf{u^2}$, thereby leaving the binding non-negativity conditions of $\mathbf{x}$ binding. Since σ_l performs row permutations and there is exactly one "1" value in each row and column, the result will satisfy the consistency conditions and it will lead to a matching. Therefore, applying an *arbitrary subset* of these disjoint cycles to $\mathbf{u^1}$ must yield a matching in $F(\mathbf{x})$.

For the other direction, let $\mathbf{u}$ be an arbitrary vertex of $F(\mathbf{x})$. The non-negativity conditions that are binding for $\mathbf{x}$ must do the same for $\mathbf{u}$. Therefore, each row of $\mathbf{u}$ belongs to one of the following two categories:

1. If there is only one non-zero element in a row of $\mathbf{x}$, then it must be 1, and $\mathbf{u}$ must be identical to both $\mathbf{u^1}$ and $\mathbf{u^2}$ in this row.

2. If there are exactly two non-zero elements in a row of $\mathbf{x}$, then $\mathbf{u^1}$ and $\mathbf{u^2}$ must have their 1 values at different positions in this row. This implies that $\mathbf{u}$ must coincide with either $\mathbf{u^1}$ or $\mathbf{u^2}$ in this row. Since the cycles in the factorization of $\Sigma_{\mathbf{u^1},\mathbf{u^2}}$ are mutually disjoint, there is a unique cycle σ_l that transforms this row of $\mathbf{u^1}$ into the corresponding row of $\mathbf{u^2}$. The preference cycle induced by σ_l has two nodes in each row and each column (where it is defined), which determines the two possible configurations of 1 values in these rows for any matching in $F(\mathbf{x})$. Therefore, $\mathbf{u}$ must coincide with the same vertex (either $\mathbf{u^1}$ or $\mathbf{u^2}$) in all of the rows involved in σ_l.

Thus, the rows $\{1, \ldots, n\}$ can be partitioned in the following way. There are cycles $\{\sigma_{l_1}, \ldots, \sigma_{l_K}\} \subseteq \{\sigma_1, \ldots, \sigma_L\}$ such that

1. $\mathbf{u}_{r\cdot} = \mathbf{u}^{\mathbf{1}}_{r\cdot} = \mathbf{u}^{\mathbf{2}}_{r\cdot}$ for all $r \notin \bigcup_l \sigma_l$,

2. $\mathbf{u}_{r\cdot} = \mathbf{u}^{\mathbf{2}}_{r\cdot} \neq \mathbf{u}^{\mathbf{1}}_{r\cdot}$ for all $r \in \bigcup_k \sigma_{l_k}$,

3. $\mathbf{u}_{r\cdot} = \mathbf{u}^{\mathbf{1}}_{r\cdot} \neq \mathbf{u}^{\mathbf{2}}_{r\cdot}$ for all $r \in \bigcup_l \sigma_l \setminus \bigcup_k \sigma_{l_k}$.

This partitioning immediately implies the relationship $\mathbf{u} = (\sigma_{l_1} \cdot \ldots \cdot \sigma_{l_K})(\mathbf{u^1})$. ■

We know that a single cycle represents an edge connecting neighboring vertices. Therefore, Lemma 3.3 implies that the set of vertices of the smallest face containing a diagonal, ordered by their relationship to one of the endpoints of the diagonal, is a lattice. Figure

3.2 illustrates this in the case in which the factorization of the permutation contains three cycles.

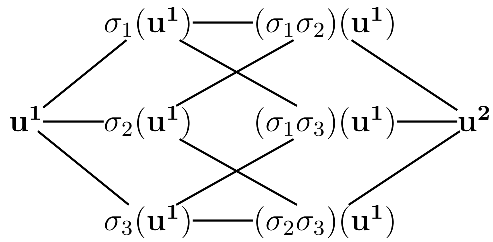


Figure 3.2: Smallest face containing a diagonal

For C-stability to be satisfied, all these vertices must be stable. Using the special structure of $F(\mathbf{x})$ and the condition obtained for neighboring vertices, we can generalize the necessary condition in the following way.

Proposition 3.6 (DC is still necessary) Assume that the fractional matching $\mathbf{x}$ lies on a line segment connecting two vertices of P', say $\mathbf{u^1}$ and $\mathbf{u^2}$. If $\mathbf{x}$ is C-stable, then every preference cycle in the graph representation that is induced by any of the row cycles in the factorization of $\Sigma_{\mathbf{u^1},\mathbf{u^2}}$ must satisfy (DC).

The following example shows that this condition is necessary, but it is still not sufficient.

Example 3.6 (DC is not sufficient) Consider the setting illustrated in Figure 3.3 (a). It is easy to verify that the matchings defined by the crossed and gray nodes are both stable.

Therefore, the following fractional matching is weakly stable:

$$\mathbf{y} = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

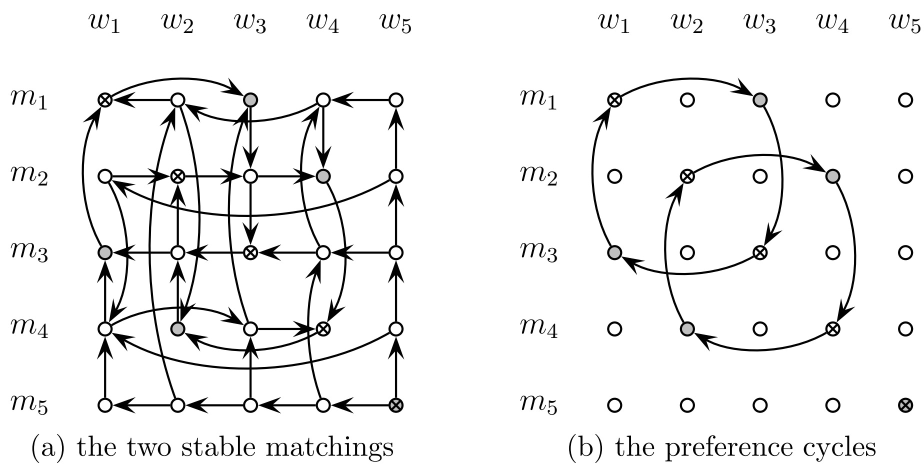


Figure 3.3: Preferences and the directed cycle condition in Example 3.6

The cycle factorization here is given by (13)(24). It can be seen in Figure 3.3 (b) that both cycles define a directed preference cycle in the graph. However, the fractional matching $\mathbf{y}$ is not C-stable since it can also be expressed as a convex combination of two unstable matchings:

$$\begin{bmatrix} 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$\square$

3.5.3 Fractional matchings lying on edges and diagonals - characterization

What causes C-stability to fail in Examples 3.5 and 3.6? Although the condition (DC) guarantees stability for pairs involved in the preference cycles, off-cycle pairs might still be able to block some of the matchings in the smallest containing face. Therefore, for a preference-based characterization, we also need conditions that rule out blocking for these off-cycle pairs.

For a given fractional matching $\mathbf{x}$ that lies on an edge or a diagonal of the matching polytope P', let $\{\rho_1, \ldots, \rho_L\}$ denote the set of preference cycles and isolated nodes defined by the pairs matched at $\mathbf{x}$. For each ρ_i, let M_{ρ_i} and W_{ρ_i} be the sets of men and women involved in ρ_i, respectively. In Example 3.5, we see a failure of stability within a group $M_{\rho_i} \times N_{\rho_i}$ and Example 3.6 illustrates a case in which stability between two groups $M_{\rho_i} \times N_{\rho_i}$ and $M_{\rho_j} \times N_{\rho_j}$ is violated. We consider these cases separately since they will lead to different stability conditions.

External stability First, consider the case in which $m \in M_{\rho_i}$ and $w \in W_{\rho_j}$ for some $i \neq j$. For an illustration, see Figure 3.4. Denote the set of partners of m and w at fractional matching $\mathbf{x}$ by $W(m)$ and $M(w)$:

- $W(m) \doteq \{w' \in W : x_{mw'} > 0\}$,
- $M(w) \doteq \{m' \in M : x_{m'w} > 0\}$.

Since $\mathbf{x}$ lies on an edge or a diagonal, there can be at most two non-zero elements in every row and column. This also means that $W(m)$ and $M(w)$ cannot have more than two elements. Let $w_-(m)$ and $w_+(m)$ denote m's worst and best partner at matching $\mathbf{x}$.[16] The partners $m_-(w)$ and $m_+(w)$ of a woman w are similarly defined.

It is easy to see that the pair (m, w) violates C-stability if and only if m is better than w's worst partner and w is better than m's worst partner: $m \succ_w m_-(w)$ and $w \succ_m w_-(m)$. In this case, there must be a matching in $F(\mathbf{x})$ where both m and w get their worst partners at $\mathbf{x}$, which creates an incentive for them to block. This argument leads us to the following condition:

(ES) A pair $(m, w) \in M_{\rho_i} \times W_{\rho_j}$ where $i \neq j$ satisfies *external stability* if either $m \prec_w m_-(w)$ or $w \prec_m w_-(m)$.

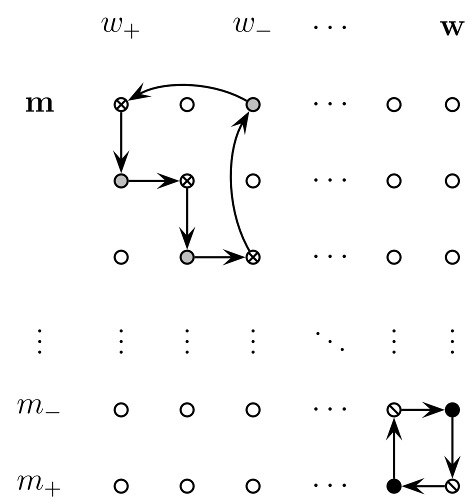
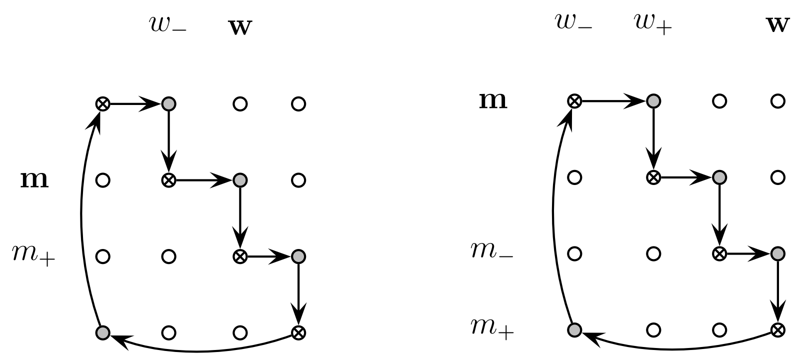


Figure 3.4: External stability

Internal stability Now take a pair (m, w) from the same group $M_{\rho_i} \times W_{\rho_i}$. We have to distinguish between two cases here:

1. Man m and woman w spend positive time together at $\mathbf{x}$: $x_{mw} > 0$, i.e., (m, w) is contained in ρ_i. (Illustrated in Figure 3.5 (a).) In this case (DC) is also sufficient. If ρ_i

[16] If m has only one partner at $\mathbf{x}$, $w_-(m) = w_+(m)$.



(a) on-cycle pairs (b) off-cycle pairs

Figure 3.5: Internal stability

is an isolated node, man m and woman w already spend the entire time together; hence, they cannot block. If ρ_i is not an isolated node and preferences form a directed cycle, then either m or w already has his or her best partner, $m = m_+(w)$ or $w = w_+(m)$ holds.

2. Man m and woman w do not spend positive time together at $\mathbf{x}$: $x_{mw} = 0$, i.e., (m, w) is not contained in ρ_i. (Illustrated in Figure 3.5 (b).) In this case, ρ_i cannot be an isolated node, and both $M(w)$ and $W(m)$ must contain two elements. We know from (DC) that the men involved in ρ_i compare the two sets of possible partners the same way and women have opposite preferences. This means that either m gets $w_+(m)$ and w gets $m_-(w)$, or m gets $w_-(m)$ and w gets $m_+(w)$. They might block if $m \succ_w m_-(w)$ and $w \succ_m w_-(m)$, but blocking happens if and only if $m \succ_w m_+(w)$ or $w \succ_m w_+(m)$. This leads to our next condition:

(IS) A pair $(m, w) \in M_{\rho_i} \times W_{\rho_i}$ such that $x_{mw} = 0$ satisfies *internal stability* if

$$[m \succ_w m_-(w) \text{ and } w \succ_m w_-(m)] \text{ implies } [m \prec_w m_+(w) \text{ and } w \prec_m w_+(m)].$$

Since we have derived stability conditions for each possible pair (m, w), we have proven the following characterization result for fractional matchings lying on diagonals.

Theorem 3.2 (characterization using preferences) Let $\mathbf{x}$ be a fractional matching that lies either on an edge or a diagonal of P'. Let $\{\rho_1, \ldots, \rho_L\}$ be the set of preference cycles and isolated nodes defined by the pairs matched at $\mathbf{x}$, and M_{ρ_i} and W_{ρ_i} be the sets of men and women involved in ρ_i, for each i. Then $\mathbf{x}$ is C-stable if and only if the following conditions are satisfied:

1. For each i, if ρ_i is a cycle, then it satisfies condition (DC);[17]

2. For each i, the pairs in $(M_{\rho_i} \times W_{\rho_i}) \setminus \rho_i$ satisfy condition (IS);

3. For each $i \neq j$, the pairs in $M_{\rho_i} \times W_{\rho_j}$ satisfy condition (ES).

3.5.4 Relationship to strong stability

We can immediately extend Theorem 3.2 to fractional matchings at which each agent spends positive time with at most two partners. Although these fractional matchings do not necessarily lie on an edge or a diagonal of P',[18] it is easy to see that the special structure lets us identify unique preference cycles and isolated nodes defined by the positive coordinates with the same properties as before. We know that a fractional matching is C-stable if and only if every vertex of the smallest containing face is stable. Since the vertices can be reconstructed

[17] We can treat isolated nodes as cycles with zero edges. In this case, they trivially satisfy condition (DC).

[18] For example, the following fractional matching has at most two positive numbers in every row and column but cannot be expressed as a convex combination of two matchings:

$$\mathbf{x} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 2/3 & 1/3 \end{bmatrix}.$$

using these cycles, Theorem 3.2 holds for all such fractional matchings as well. This observation can be used to discuss the difference between C-stability and strong stability since in a strongly stable matching everyone can have at most two partners.

Remember that in a strongly stable matching, at least one member of each pair (m, w) cannot spend time with someone less preferred than the other. If we compare this to Theorem 3.2, we can see that this is exactly what conditions (ES) and (DC) require. However, the condition (IS) is weaker. Pairs who belong to the same group $M_{\rho_i} \times W_{\rho_i}$ can both spend positive time with someone less desired than the other if the cycle ρ_i does not let this happen simultaneously. This shows that weakening strong stability extends the set of stable matching even among matchings in which no one spends time with more than two partners.

The difference between the two stability concepts is more significant for matchings in which someone is paired with more than two partners over time. As we have seen, no such fractional matching can be strongly stable, but C-stability can still be satisfied. In the next subsection, we will take the results of Theorem 3.2 one step further to characterize C-stability in the general case.

3.5.5 Generalization

We can use the results we have derived so far to characterize general C-stable matchings. First, notice that based on a given fractional matching, the whole market (M, W) can be decomposed into a collection of submarkets such that people from different submarkets do not spend any time together, and this decomposition is the finest possible.[19] Moreover, every submarket must contain the same number of men and women since the total time spent

[19]The following algorithm can find this decomposition $(M_k, W_k)_k$. Take the first available man, m_1. Find all his partners, $W(m_1)$, then all the partners of the women in $W(m_1)$, $M(W(m_1))$, and so on. Continue this until the sets stop changing between iterations. The result will be the first submarket $(M_1 \times W_1)$. Then keep repeating the process with the remaining agents until there is someone left. It is straightforward to verify that this algorithm will lead to the collection of submarkets with the desired properties.

with the other side of the submarket must add up to the same length for the two groups. Given such a decomposition, a fractional matching $\mathbf{x}$ defined for the whole market defines a fractional matching for each submarket. Therefore, we can use Theorem 3.2 to exlude blocking pairs whose members belong to the same submarket and the external stability condition to exclude blocking pairs whose members are from different submarkets.

Theorem 3.3 (General characterization) Given a fractional matching $\mathbf{x}$, let $(M_k, W_k)_{k=1}^K$ be the decomposition of the original market into submarkets as described above. Then $\mathbf{x}$ is C-stable if and only if the following holds.

1. Condition (ES) holds for every pair $(m, w) \in M_k \times W_{k'}$, for every $k \neq k'$.
2. For any submarket $M_k \times W_k$ and for any set of preference cycles and nodes $\{\rho_1, \dots, \rho_L\}$ defined on $M_k \times W_k$ such that:

   (a) $\mathbf{x}$ assigns positive time to every node in the ρ_i's,

   (b) the ρ_i's induce a partition of men and women in $M_k \times W_k$,

   the conditions of Theorem 3.2 must be true.

Proof. Note that the decomposition of the market also implies that every matching in $F(\mathbf{x})$ can be decomposed in a similar way. For every matching $\mathbf{y} \in F(\mathbf{x})$,

- the submatrix of $\boldsymbol{y}$ that corresponds to $M_k \times W_k$ defines a matching in $M_k \times W_k$;
- the submatrix of $\boldsymbol{y}$ that corresponds to $M_k \times W_{k'}$ contains only 0 elements.

Therefore, a matching in $F(\mathbf{x})$ is stable if and only if its restriction to any submarket $M_k \times W_k$ is a stable matching, and no pair in $M_k \times W_{k'}$, where $k \neq k'$ has an incentive to block it.

1. Take a pair $(m, w) \in M_k \times W_{k'}$. Since the two agents belong to different submarkets, there must be a matching in $F(\mathbf{x})$ at which both of them are matched to their least preferred partners at $\mathbf{x}$. If the pair (m, w) blocks some matching in $F(\mathbf{x})$, then they will definitely block the ones at which they are both matched to their least preferred partners at $\mathbf{x}$. Therefore, this pair cannot block any matching in $F(\mathbf{x})$ if and only if condition (ES) is satisfied.

2. Consider the submarket $M_k \times W_k$. Let $\rho_1, \ldots, \rho_L$ have the above described properties. We will construct all of the matchings that are defined in this submarket by these preference cycles and isolated nodes. Remember that we can assign the 1-elements in two ways for each preference cycle, and we have only one (trivial) option to do that for the degenerate cycles. No matter how we assign the 1 elements in each preference cycle, property (b) implies that there will be exactly one 1-element in each row and column of the block in the the matrix. Therefore, this process will lead to a matching in the submarket $M_k \times W_k$. On the other hand, property (a) guarantees that all the inequalities that are binding for $\mathbf{x}$ for $M_k \times W_k$ remain binding, and the preference cycles will correspond to some edges connecting vertices in the matching polytope defined by market $M_k \times W_k$. We can apply the former necessary and sufficient conditions. Moreover, repeating the same steps for every such set of preference cycles will produce all the matchings in $M_k \times W_k$ that are consistent with matchings in $F(\mathbf{x})$. ■

If $\mathbf{x}$ is a matching, only isolated points have positive weights. Therefore, only the external stability condition is required which coincides with the usual stability condition in this case. This demonstrates how C-stability generalizes the original stability concept from single matchings to entire faces of the polytope P'.

3.6 Conclusions

This chapter has presented an alternative stability concept for fractional matchings in marriage markets with time dimension. The new definition, C-stability, is less restrictive than strong stability proposed by Roth et al. (1993). C-stability is by definition equivalent to the non-existence of unstable implementing time schedules. We have characterized C-stability both geometrically and in terms of the underlying preferences.

We have shown that a fractional matching $\mathbf{x}$ is C-stable if and only if the smallest face of the matching polytope that contains it, $F(\mathbf{x})$, has only stable vertices. Moreover, the set of C-stable matchings is the union of the faces shared by the stable matching and the matching polytopes.

Using these geometric results, we have obtained conditions on the underlying preferences that are necessary and sufficient for C-stability. The general theorem characterizes C-stability of a fractional matching $\mathbf{x}$ using faces generated by line segments in $F(\mathbf{x})$. Pairs that are matched at some of the vertices of this generated face define cycles and isolated nodes in the preference graph such that they partition the sets of men and women. We have derived conditions that must hold for such a construction. First, all such cycles in the preference graph must be directed. Second, the stability of an off-cycle pair depends on whether or not the man and woman belong to groups defined by the same cycle. If they do not, only one of them can have a less preferred partner on the cycle. If they do, stability is less restrictive. They can both have less preferred partners on the cycle as long as they both have more preferred partners as well since they cannot be matched with their worst options simultaneously. This last condition highlights the difference between strong stability and our concept.

We used two assumptions in our model. First, we had an equal number of men and women, and second, each pair was mutually acceptable. It would be interesting to know how crucial these assumptions are for our results.

Generalizing the first should be straightforward. It is known that, given the preferences, the set of matched agents is always the same in every stable matching.[20] The agents who are always single in stable matchings would act as redundant zero rows or columns in our matrices and would not change our results.

The second assumption should not be crucial either. The polytopes are still integral even if we allow for unacceptable pairs to exists.[21] Therefore, slightly modified versions of the theorems that take unacceptable pairs into account would most likely still work.

There are many directions for further research. Since the general condition still has combinatorial features, checking it could be tiresome when agents have more than two partners in a fractional matching. It would be useful to know whether the general condition can be simplified in this case.

Conducting a similar analysis for different types of matching problems (e.g., the roommate problem) would also be of considerable interest. In such settings, however, other complications may arise since even the basic results such as the existence of stable matchings or the integrability of the matching polytope are not guaranteed to hold.

[20] See e.g., Roth and Sotomayor (1992).
[21] See Roth et al. (1993).

Bibliography

Abeledo, H. G., Blum, Y., and Rothblum, U. G. (1996). Canonical monotone decompositions of fractional stable matchings. *International Journal of Game Theory*, 25(2):161–176.

Allison, J. R. (1990). Five ways to keep disputes out of court. *Harvard Business Review*, 68(1):166–8.

Alonso, R. and Câmara, O. (2016). Persuading voters. *American Economic Review*, 106(11):3590–3605.

Balinski, M. and Ratier, G. (1997). Of stable marriages and graphs, and strategy and polytopes. *SIAM review*, 39(4):575–604.

Board, S. (2009). Revealing information in auctions: the allocation effect. *Economic Theory*, 38(1):125–135.

Brooks, R. R., Landeo, C. M., and Spier, K. E. (2010). Trigger happy or gun shy? Dissolving common-value partnerships with Texas shootouts. *The RAND Journal of Economics*, 41(4):649–673.

Brown, A. L. and Velez, R. A. (2016). The costs and benefits of symmetry in common-ownership allocation problems. *Games and Economic Behavior*, 96:115–131.

Condorelli, D. and Szentes, B. (2017). Information design in the hold-up problem. Manuscript submitted for publication.

Cramton, P., Gibbons, R., and Klemperer, P. (1987). Dissolving a partnership efficiently. *Econometrica: Journal of the Econometric Society*, pages 615–632.

Cremer, J. and McLean, R. P. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257.

de Frutos, M. A. (2000). Asymmetric price-benefits auctions. *Games and Economic Behavior*, 33(1):48–71.

de Frutos, M.-A. and Kittsteiner, T. (2008). Efficient partnership dissolution under buy-sell clauses. *The RAND Journal of Economics*, 39(1):184–198.

Echenique, F., Lee, S., Shum, M., and Yenmez, M. B. (2013). The revealed preference theory of stable and extremal stable matchings. *Econometrica*, 81(1):153–171.

Ensthaler, L., Giebe, T., and Li, J. (2014). Speculative partnership dissolution with auctions. *Review of Economic Design*, 18(2):127–150.

Eső, P. and Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731.

Fieseler, K., Kittsteiner, T., and Moldovanu, B. (2003). Partnerships, lemons, and efficient trade. *Journal of Economic Theory*, 113(2):223–234.

Fu, H., Jordan, P., Mahdian, M., Nadav, U., Talgam-Cohen, I., and Vassilvitskii, S. (2012). Ad auctions with data. In *Algorithmic Game Theory*, pages 168–179. Springer.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719.

Gusfield, D. and Irving, R. W. (1989). *The stable marriage problem: structure and algorithms*, volume 54. MIT press Cambridge.

Hart, O. and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158.

Hauswald, R. B. and Hege, U. (2009). Ownership and control in joint ventures: Theory and evidence. Mimeo.

Jehiel, P. and Pauzner, A. (2006). Partnership dissolution with interdependent values. *The RAND Journal of Economics*, 37(1):1–22.

Jiang, Y. and Guo, H. (2015). Design of consumer review systems and product pricing. *Information Systems Research*, 26(4):714–730.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kesten, O. and Ünver, M. U. (2015). A theory of school-choice lotteries. *Theoretical Economics*, 10(2):543–595.

Kittsteiner, T. (2003). Partnerships and double auctions with interdependent valuations. *Games and Economic Behavior*, 44(1):54–76.

Kittsteiner, T., Ockenfels, A., and Trhal, N. (2012). Partnership dissolution mechanisms in the laboratory. *Economics Letters*, 117(2):394–396.

Koessler, F. and Renault, R. (2012). When does a firm disclose product information? *The RAND Journal of Economics*, 43(4):630–649.

Koessler, F. and Skreta, V. (2016). Informed seller with taste heterogeneity. *Journal of Economic Theory*, 165:456–471.

Krishna, V. and Perry, M. (1998). Efficient mechanism design. Mimeo, Penn State University.

Kwark, Y., Chen, J., and Raghunathan, S. (2014). Online product reviews: Implications for retailers and competing manufacturers. *Information systems research*, 25(1):93–110.

Landeo, C. M. and Spier, K. E. (2013). Shotgun mechanisms for common-value partnerships: The unassigned-offeror problem. *Economics Letters*, 121(3):390–394.

Lauermann, S. and Nöldeke, G. (2014). Stable marriages and search frictions. *Journal of Economic Theory*, 151:163–195.

Levin, J. and Tadelis, S. (2005). Profit sharing and the role of professional partnerships. *The Quarterly Journal of Economics*, 120(1):131–171.

Lewis, T. R. and Sappington, D. E. (1989). Countervailing incentives in agency problems. *Journal of Economic Theory*, 49(2):294–313.

Lewis, T. R. and Sappington, D. E. (1994). Supplying information to facilitate price discrimination. *International Economic Review*, 35(2):309–327.

Li, H. and Shi, X. (2017a). Discriminatory information disclosure. *American Economic Review*, 107(11):3363–85.

Li, H. and Shi, X. (2017b). Optimal discriminatory disclosure. Working Paper, Vancouver School of Economics, University of British Columbia.

Li, X., Hitt, L. M., and Zhang, Z. J. (2011). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, 27(4):9–42.

Lipsky, D. B. and Seeber, R. L. (1998). The appropriate resolution of corporate disputes: A report on the growing use of ADR by US corporations. *Martin and Laurie Scheinman Institute on Conflict Resolution*, page 4.

Loertscher, S. and Wasser, C. (2016). Optimal structure and dissolution of partnerships. Mimeo.

McAfee, R. P. (1992). Amicable divorce: Dissolving a partnership with simple mechanisms. *Journal of Economic Theory*, 56(2):266–293.

Mezzetti, C. (2004). Mechanism design with interdependent valuations: Efficiency. *Econometrica*, 72(5):1617–1626.

Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, 20(5):1089–1122.

Moldovanu, B. (2002). How to dissolve a partnership. *Journal of Institutional and Theoretical Economics JITE*, 158(1):66–80.

Myerson, R. B. and Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281.

Neeman, Z. (1999). Property rights and efficiency of voluntary bargaining under asymmetric information. *The Review of Economic Studies*, 66(3):679–691.

Ornelas, E. and Turner, J. L. (2007). Efficient dissolution of partnerships and the structure of control. *Games and Economic Behavior*, 60(1):187–199.

Ottaviani, M. and Prat, A. (2001). The value of public information in monopoly. *Econometrica*, 69(6):1673–1683.

Roesler, A.-K. and Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–2080.

Roth, A. E., Rothblum, U. G., and Vande Vate, J. H. (1993). Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research*, 18(4):803–828.

Roth, A. E. and Sotomayor, M. A. O. (1992). *Two-sided matching: A study in game-theoretic modeling and analysis*. Number 18. Cambridge University Press.

Rothblum, U. G. (1992). Characterization of stable matchings as extreme points of a polytope. *Mathematical Programming*, 54(1-3):57–67.

Salant, Y. and Siegel, R. (2016). Reallocation costs and efficiency. *American Economic Journal: Microeconomics*, 8(1):203–227.

Schweizer, U. (2006). Universal possibility and impossibility results. *Games and Economic Behavior*, 57(1):73–85.

Segal, I. and Whinston, M. D. (2011). A simple status quo that ensures participation (with application to efficient bargaining). *Theoretical Economics*, 6(1):109–125.

Shapley, L. S. and Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory*, 1(1):111–130.

Shontz, D., Kipperman, F., and Soma, V. (2011). Business-to-Business Arbitration in the United States.

Teo, C.-P. and Sethuraman, J. (1998). The geometry of fractional stable matchings and its applications. *Mathematics of Operations Research*, 23(4):874–891.

Turner, J. L. (2013). Dissolving (in) effective partnerships. *Social Choice and Welfare*, 41(2):321–335.

Van Essen, M. and Wooders, J. (2016). Dissolving a partnership dynamically. *Journal of Economic Theory*, 166:212–241.

Vande Vate, J. H. (1989). Linear programming brings marital bliss. *Operations Research Letters*, 8(3):147–153.

Veugelers, R. and Kesteloot, K. (1996). Bargained shares in joint ventures among asymmetric partners: Is the matthew effect catalyzing? *Journal of Economics*, 64(1):23–51.

Wasser, C. (2013). Bilateral k+1-price auctions with asymmetric shares and values. *Games and Economic Behavior*, 82:350–368.

Williams, S. R. (1999). A characterization of efficient, Bayesian incentive compatible mechanisms. *Economic Theory*, 14(1):155–180.

Yamashita, T. (2018). Optimal public information disclosure by mechanism designer. Working Paper, Toulouse University.

Zhang, T., Li, G., Cheng, T., and Lai, K. K. (2017). Welfare economics of review information: Implications for the online selling platform owner. *International Journal of Production Economics*, 184:69–79.

Appendix A

Appendix for Chapter 1

A.1 Proof of Proposition 1.3

Proposition 1.2 identifies a finite set of posterior beliefs that can be used to construct the optimal disclosure policy. Therefore, we can replace the maximization problem (1.1) by the following linear programming problem:

$$\max \quad \sum_{k=1}^{J} \lambda_k \Pi(\boldsymbol{e^{(k)}}; \alpha) + \sum_{k_1=1}^{J} \sum_{k_2=J+1}^{K} \eta_{k_1,k_2} \Pi(\boldsymbol{b^{(k_1,k_2)}}; \alpha) \tag{A.1}$$

$$\text{s.t.} \quad \sum_{k=1}^{J} \lambda_k \boldsymbol{e^{(k)}} + \sum_{k_1=1}^{J} \sum_{k_2=J+1}^{K} \eta_{k_1,k_2} \boldsymbol{b^{(k_1,k_2)}} = \boldsymbol{\rho} \tag{A.2}$$

$$\sum_{k=1}^{J} \lambda_k + \sum_{k_1=1}^{J} \sum_{k_2=J+1}^{K} \eta_{k_1,k_2} = 1 \tag{A.3}$$

$$\lambda_k, \eta_{k_1,k_2} \geqq 0 \quad \forall k, k_1, k_2$$

This linear programming problem can be further simplified as follows. First, constraint (A.3) is redundant since it is the sum of all the linear equations contained in constraint (A.2).

Second, note that the following equalities hold for every $\psi_k, \psi_{k_1} \in \Psi^{(1)}$ and $\psi_{k_2} \in \Psi^{(2)}$:

$$\begin{aligned}
\Pi(\boldsymbol{b^{(k_1,k_2)}};\alpha) &= \langle \boldsymbol{b^{(k_1,k_2)}}, \boldsymbol{v_{2\cdot}}\rangle; \\
\Pi(\boldsymbol{e^{(k)}};\alpha) &= \max\{\alpha\langle \boldsymbol{e^{(k)}}, \boldsymbol{v_{1\cdot}}\rangle, \langle \boldsymbol{e^{(k)}}, \boldsymbol{v_{2\cdot}}\rangle\} = \max\{\alpha v_{1k}, v_{2k}\}.
\end{aligned}$$

Therefore, we can simplify the objective function by subtracting from it the constant

$$\langle \boldsymbol{\rho}, \boldsymbol{v_{2\cdot}}\rangle = \sum_{k=1}^{J} \lambda_k \langle \boldsymbol{e^{(k)}}, \boldsymbol{v_{2\cdot}}\rangle + \sum_{k_1=1}^{J}\sum_{k_2=J+1}^{K} \eta_{k_1,k_2}\langle \boldsymbol{b^{(k_1,k_2)}}, \boldsymbol{v_{2\cdot}}\rangle.$$

To ease notation, assume without loss of generality that the states are labeled such that the seller's net profit from restricting the output and selling only to type 1, $\alpha v_{1k} - v_{2k}$, is positive exactly in the first $H \leqq J$ states.

Using the observations made above, we can rewrite linear programming problem (A.1) as follows:

$$\begin{aligned}
\max \quad & \sum_{k=1}^{H} \lambda_k(\alpha v_{1k} - v_{2k}) && \text{(A.4)}\\
\text{s.t.} \quad & \sum_{k=1}^{H} \lambda_k \boldsymbol{e^{(k)}} + \sum_{k=H+1}^{J} \lambda_k \boldsymbol{e^{(k)}} + \sum_{k_1=1}^{J}\sum_{k_2=J+1}^{K} \eta_{k_1,k_2} \boldsymbol{b^{(k_1,k_2)}} = \boldsymbol{\rho} && \text{(A.5)}\\
& \lambda_k, \eta_{k_1,k_2} \geqq 0 \qquad \text{for every } k, k_1, k_2
\end{aligned}$$

Linear programming problem (A.4) shows the importance of the first H states in the seller's optimal information design problem in two ways. First, if we fix the conditional distribution of the first H fully revealing posteriors, the seller has an incentive to assign as much probability as possible to reaching this set. Second, if we keep the overall probability assigned to the first H posteriors fixed, the seller is interested in shifting probabilities toward states with higher net gains.

If types are ranked in the same way for every state then $J = K$, and no $\boldsymbol{b}^{(k_1,k_2)}$ vector is contained in problem (A.4). The only feasible solution (and hence the only optimal solution) must be $\boldsymbol{\lambda}^* = \boldsymbol{\rho}$, which trivially satisfies the properties in the statement.

If types are not ranked uniformly, then $J < K$. In this case, there are $J(K-J)$ vectors of the form $\boldsymbol{b}^{(k_1,k_2)}$ in problem (A.4). Assume to the contrary that there exist $k, k' \leqq H$ such that the following inequalities hold:

$$\alpha v_{1k} - v_{2k} > \alpha v_{1k'} - v_{2k'};$$
$$\lambda_k^* < \rho_k;$$
$$\lambda_{k'}^* > 0.$$

Consider the k-th linear equation in (A.5). Since $\lambda_k^* < \rho_k$, there must exist $l > J$ such that $\eta_{k,l}^* > 0$. Now let $\varepsilon \doteq \min\{\eta_{k,l}^*, \rho_k - \lambda_k^*, \rho_{k'} - \eta_{k',l}^*, \lambda_{k'}^*\} > 0$ and consider the solution that is identical to $(\boldsymbol{\lambda}^*, \boldsymbol{\eta}^*)$ except that the values $\lambda_k^*, \lambda_{k'}^*, \eta_{k,l}^*$, and $\eta_{k',l}^*$ are replaced by $\lambda_k^* + \varepsilon, \lambda_{k'}^* - \varepsilon, \eta_{k,l}^* - \varepsilon$, and $\eta_{k',l}^* + \varepsilon$, respectively. The k-th and k'-th linear equations and the non-negativity conditions are still satisfied, and no other equations in (A.5) are affected. This change, however, increases the value of the objective function by the amount $\varepsilon((\alpha v_{1k} - v_{2k}) - (\alpha v_{1k'} - v_{2k'})) > 0$, contradicting the optimality of $(\boldsymbol{\lambda}^*, \boldsymbol{\eta}^*)$.

Therefore, there must exist a threshold $T > 0$ such that for each $k \leqq H$,

$$\lambda_k^* = \begin{cases} \rho_k & \text{if } \alpha v_{1k} - v_{2k} > T; \\ 0 & \text{if } \alpha v_{1k} - v_{2k} < T. \end{cases}$$

The same reasoning applies if there exists a $k \leqq H$ such that $\lambda_k^* < \rho_k$ and an l such that $H < l \leqq J$ and $\lambda_l^* > 0$. ■

A.2 Proof of Corollary 1.1

Start out with the prior $\boldsymbol{\rho}$ defined on state space Ψ, and an optimal distribution as described in Proposition 1.2. Let $\boldsymbol{\lambda} \in \mathbb{R}^J$ denote the probabilities assigned to posteriors $\boldsymbol{e^{(k)}}$, where $k = 1, \ldots, J$, and $\boldsymbol{\eta} \in \mathbb{R}^{J\times(K-J)}$ denote the probabilities assigned to posteriors $\boldsymbol{b^{(k_1,k_2)}}$, where $k_1 = 1, \ldots, J$ and $k_2 = J+1, \ldots, K$. We will show that the optimal posteriors and the prior belief can be equivalently represented in the decomposed state space such that the assumptions of the corollary hold and the new distribution satisfies the law of total probability.

First, note that a belief on $\Xi \times \Omega^{(1)} \times \Omega^{(2)}$ can be defined by a three-dimensional array $\boldsymbol{\mu} \in \mathbb{R}^{2\times J\times(K-J)}$ such that μ_{jkl} represents the probability of state $(\xi_j, \omega_k^{(1)}, \omega_{J+l}^{(2)})$ for every $j = 1, 2$, $k = 1 \ldots, J$, and $l = 1, \ldots, K-J$. Then the posterior beliefs on Ψ in the optimal distribution above can be represented in the decomposed state space as follows:

- For every $k' = 1, \ldots, J$, the fully revealing posterior $\boldsymbol{e^{(k')}} \in \mathbb{R}^K$ is represented by $\boldsymbol{\mu^{(k')}}$ such that $\mu_{1k'1}^{(k')} = 1$ and $\mu_{jkl}^{(k')} = 0$ for each $(j, k, l) \neq (1, k', 1)$,
- For every $k_1 = 1, \ldots, J$ and $k_2 = J+1, \ldots, K$, the posterior $\boldsymbol{b^{(k_1,k_2)}}$ is equivalent to $\boldsymbol{\nu^{(k_1,k_2)}}$ such that $\nu_{1k_1k_2}^{(k_1,k_2)} = b_{k_1}^{(k_1,k_2)}$, $\nu_{2k_1k_2}^{(k_1,k_2)} = b_{k_2}^{(k_1,k_2)}$, and $\nu_{jkl}^{(k_1,k_2)} = 0$ for every $(j, k, l) \notin \{(1, k_1, k_2), (2, k_1, k_2)\}$.

Notice that the vertical state variable is fully revealed for both groups of beliefs, while the horizontal state variable is fully revealed only for the first group. This distribution trivially satisfies the law of total probability if we define the prior $\boldsymbol{\nu}$ as

$$\boldsymbol{\nu} \doteq \sum_{k'=1}^{J} \lambda_{k'} \boldsymbol{\mu^{(k')}} + \sum_{k_1=1}^{J} \sum_{k_2=J+1}^{K} \eta_{k_1,k_2} \boldsymbol{\nu^{(k_1,k_2)}}.$$

Since the problems defined in the original and in the decomposed state space are essentially the same,[1] the distribution of posteriors defined over the decomposed state space must be optimal as well. ■

A.3 Proof of Proposition 1.5

Part 1. The proof is essentially identical to that of the first part of Proposition 1.1. Consider an arbitrary posterior $\tilde{\boldsymbol{\rho}}$. Assume that it is optimal to sell to types $\theta_1, \ldots, \theta_j$ at this posterior. Since types are ranked uniformly, it follows that by disclosing all the information and setting the price equal to θ_j's valuation, the seller can always sell to at least the same set of types; and hence, she can obtain at least the same profit in expectation as she did at posterior $\tilde{\boldsymbol{\rho}}$. Therefore, the optimal profit under full disclosure must be at least the same as the profit obtained at posterior $\tilde{\boldsymbol{\rho}}$.

Given any distribution of posteriors consistent with the law of total probability, we can replace each posterior with its decomposition into a distribution of fully disclosing posteriors without the seller losing expected profit. Therefore, full disclosure must be optimal.

Part 2. Assume without loss of generality that types 1 and 2 are not ranked uniformly for every state: there are states ψ_{k_1} and ψ_{k_2} such that $v_{1k_1} > v_{2k_1}$ and $v_{1k_2} < v_{2k_2}$. Since expected valuations are linear in the probabilities, there exists a unique γ^* such that expected valuations are equalized at the posterior $\boldsymbol{\rho}^* \doteq \gamma^* \boldsymbol{e}^{(\boldsymbol{k_1})} + (1-\gamma^*)\boldsymbol{e}^{(\boldsymbol{k_2})}$, i.e., $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle = \langle \boldsymbol{\rho}^*, \boldsymbol{v_{2\cdot}} \rangle$. We will now show that there is a nonempty, open set of type distributions such that the seller prefers disclosing no information to disclosing all information at posterior $\boldsymbol{\rho}^*$. Therefore, if a fully revealing signal structure is used for any prior $\boldsymbol{\rho} \in \operatorname{int} \Delta^{K-1}$, then the seller can

[1] We split up the states in Ψ by the realization of one of the vertical state variables; hence we do not change the distribution of the valuation profiles.

always increase her revenue by reallocating some probability from $\boldsymbol{e}^{(\boldsymbol{k_1})}$ and $\boldsymbol{e}^{(\boldsymbol{k_2})}$ to $\boldsymbol{\rho}^*$ in the distribution of posteriors.

We can guarantee that the optimal price at $\boldsymbol{\rho}^*$ is equal to the valuation of types 1 and 2 by assigning sufficiently high probability to these two types. First, the seller does not want to charge a higher price than $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle = \langle \boldsymbol{\rho}^*, \boldsymbol{v_{2\cdot}} \rangle$ if the following condition holds:

$$\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle \, \alpha\Big(\Big\{\theta_i \colon \langle \boldsymbol{\rho}^*, \boldsymbol{v_{i\cdot}} \rangle \geqq \langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle\Big\}\Big) > \langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle \, \alpha\Big(\Big\{\theta_i \colon \langle \boldsymbol{\rho}^*, \boldsymbol{v_{i\cdot}} \rangle \geqq \langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle\Big\}\Big) \tag{A.6}$$

for each θ_j such that $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle > \langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle$. This condition is definitely satisfied if

$$\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle > \langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle \, (1 - \alpha_1 - \alpha_2)$$

is true for every θ_j such that $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle > \langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle$. This leads to the sufficient condition

$$\alpha_1 + \alpha_2 > 1 - \frac{\langle \boldsymbol{\rho}, \boldsymbol{v_{1\cdot}} \rangle}{\max_{\theta_j} \langle \boldsymbol{\rho}, \boldsymbol{v_{j\cdot}} \rangle} \in [0, 1). \tag{A.7}$$

Similarly, the seller does not prefer to charge a price lower than $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle = \langle \boldsymbol{\rho}^*, \boldsymbol{v_{2\cdot}} \rangle$ if (A.6) holds for each θ_j such that $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle < \langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle$. This condition is satisfied if the inequality

$$\langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle (\alpha_1 + \alpha_2) > \langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle$$

is true for every θ_j such that $\langle \boldsymbol{\rho}^*, \boldsymbol{v_{j\cdot}} \rangle < \langle \boldsymbol{\rho}^*, \boldsymbol{v_{1\cdot}} \rangle$. This leads to the sufficient condition

$$\alpha_1 + \alpha_2 > \frac{\sup_{\theta_j} \{\langle \boldsymbol{\rho}, \boldsymbol{v_{j\cdot}} \rangle \colon \langle \boldsymbol{\rho}, \boldsymbol{v_{j\cdot}} \rangle < \langle \boldsymbol{\rho}, \boldsymbol{v_{1\cdot}} \rangle\}}{\langle \boldsymbol{\rho}, \boldsymbol{v_{1\cdot}} \rangle} \in [-\infty, 1). \tag{A.8}$$

Hence, if $\alpha_1 + \alpha_2$ is higher than both of the thresholds in conditions (A.7) and (A.8), the optimal price at $\boldsymbol{\rho}^*$ coincides with the expected valuation of types 1 and 2. Therefore, all the surplus of types 1 and 2 is captured by the seller at $\boldsymbol{\rho}^*$. If the seller replaced this posterior with its decomposition into a convex combination of posteriors $\boldsymbol{e}^{(k_1)}$ and $\boldsymbol{e}^{(k_2)}$, she would definitely lose some expected profit on these two types since they are strictly ranked in states ψ_{k_1} and ψ_{k_2}. This loss is bounded from below by the following expression:

$$\begin{aligned} L(\alpha_1,\alpha_2) &\doteq \gamma^* \min\{\alpha_1(v_{1k_1} - v_{2k_1}), \alpha_2 v_{2k_1}\} + (1-\gamma^*)\min\{\alpha_2(v_{2k_2} - v_{1k_2}), \alpha_1 v_{1k_2}\} \\ &= (\alpha_1+\alpha_2)\Bigg(\gamma^* \min\left\{\frac{\alpha_1}{\alpha_1+\alpha_2}(v_{1k_1} - v_{2k_1}), \frac{\alpha_2}{\alpha_1+\alpha_2} v_{2k_1}\right\} \\ &\qquad + (1-\gamma^*)\min\left\{\frac{\alpha_2}{\alpha_1+\alpha_2}(v_{2k_2} - v_{1k_2}), \frac{\alpha_1}{\alpha_1+\alpha_2} v_{1k_2}\right\}\Bigg) > 0. \end{aligned}$$

By revealing all information at $\boldsymbol{\rho}^*$, the seller might also gain revenue on some other types that are buying the good at the fully revealing posteriors. However, this gain must be bounded from above by

$$G(\alpha_1,\alpha_2) \doteq (1-\alpha_1-\alpha_2)\left(\gamma^* \max_{\theta_i} v_{ik_1} + (1-\gamma^*)\max_{\theta_i} v_{ik_2}\right) > 0.$$

Starting with a pair $\alpha_1, \alpha_2 > 0$, which satisfies $\alpha_1 + \alpha_2 < 1$ and conditions (A.7) and (A.8), we can change the probabilities such that the inequality

$$L(\alpha_1,\alpha_2) > G(\alpha_1,\alpha_2) \tag{A.9}$$

is also satisfied, which guarantees the suboptimality of full disclosure at $\boldsymbol{\rho}^*$. In order to achieve this, increase the sum $\alpha_1 + \alpha_2$ such that it stays below 1 and the ratio of α_1 to α_2

is fixed. As $\alpha_1 + \alpha_2$ approaches 1, the lower bound on loss $L(\alpha_1, \alpha_2)$ strictly increases while the upper bound on gain $G(\alpha_1, \alpha_2)$ approaches zero.

The probabilities $\alpha_1, \alpha_2 > 0$, obtained in this way satisfy conditions $\alpha_1 + \alpha_2 < 1$, (A.7), (A.8), and (A.9) regardless of the probabilities assigned to the rest of the types. Since these conditions are strict inequalities, we can find a non-empty open set of type distributions that guarantee the suboptimality of full disclosure at $\boldsymbol{\rho}^*$. ■

Appendix B

Appendix for Chapter 2

B.1 Derivation of (2.3) and (2.4)

The derivation of (2.3) follows from the fact that

$$
\begin{aligned}
&\int_0^1 \int_0^{\theta_i \wedge k} \theta_i - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) \,\mathrm{d}F(\theta_i) \\
&= \int_k^1 \int_0^k \theta_i - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) \,\mathrm{d}F(\theta_i) + \int_0^k \int_0^{\theta_i} \theta_i - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) \,\mathrm{d}F(\theta_i) \\
&= \int_k^1 F(k)(\theta_i - k) + \int_0^k F(\theta_{-i}) \,\mathrm{d}\theta_{-i} \,\mathrm{d}F(\theta_i) + \int_0^k \int_0^{\theta_i} F(\theta_{-i}) \,\mathrm{d}\theta_{-i} \,\mathrm{d}F(\theta_i) \\
&= F(k) \int_k^1 1 - F(\theta)\mathrm{d}\theta + (1 - F(k)) \int_0^k F(\theta) \,\mathrm{d}\theta + F(k) \int_0^k F(\theta) \,\mathrm{d}\theta - \int_0^k F(\theta)^2 \,\mathrm{d}\theta \\
&= F(k) \int_k^1 1 - F(\theta)\mathrm{d}\theta + \int_0^k F(\theta) \left(1 - F(\theta)\right) \,\mathrm{d}\theta,
\end{aligned}
$$

where the second and third equalities follows from integration by parts.

Next, it can easily be verified that under the mechanism Γ^*,

$$\begin{aligned} U_i(\theta_i^*(r_i;k); r_i, k) &= \int_0^{\theta_i^*(r_i;k)} \theta_i^*(r_i;k) - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) - r_i(\theta_i^*(r_i;k) - k) \\ &= \theta_i^*(r_i;k) F(\theta_i^*(r_i;k)) - \int_0^{\theta_i^*(r_i;k)} \theta_{-i} \,\mathrm{d}F(\theta_{-i}) - r_i(\theta_i^*(r_i;k) - k) \\ &= \theta_i^*(r_i;k) F(\theta_i^*(r_i;k)) - \theta_i^*(r_i;k) F(\theta_i^*(r_i;k)) \\ &\qquad + \int_0^{\theta_i^*(r_i;k)} F(\theta_{-i}) \,\mathrm{d}\theta_{-i} - r_i(\theta_i^*(r_i;k) - k) \\ &= \int_0^{\theta_i^*(r_i;k)} F(\theta_{-i}) \,\mathrm{d}\theta_{-i} - r_i(\theta_i^*(r_i;k) - k), \end{aligned}$$

where the third equality follows from integration by parts.

Since $\mathcal{L}(r_1, r_2, k) = U_1(\theta_1^*(r_1,k); r_1, k) + U_2(\theta_2^*(r_2,k); r_2, k)$, (2.4) follows immediately. ∎

B.2 Proof of Lemma 2.1

We will show that truth telling is an ex-post equilibrium in the direct revelation game defined by the efficient allocation rule (2.1) and the payment rules (2.2), and that the equilibrium payoffs in this game are always non-negative. Consequently, the interim conditions (IC) and (IR) are also satisfied and truth-telling constitutes a Bayes Nash Equilibrium in the direct revelation game defined at the interim stage by the same allocation and payment rules.

To show that truth-telling is an ex-post equilibrium, we need to prove that it is always optimal for player i to report her type truthfully, knowing $-i$'s report and the fact that $-i$ follows the same strategy. Let θ_i and θ_{-i} denote the true types and θ_i' and θ_{-i}' the types reported by the players. The ex-post net utility values of player i conditional on the reports θ_i' and $\theta_{-i}' = \theta_{-i}$, defined by $u_i(\theta; q^*(\theta_1', \theta_2')) + t_i^*(\theta_1', \theta_2') - r_i(\theta_1 + \theta_2 - k)$, are displayed in Table B.1. Comparing the net payoff terms given $-i$'s reported true type, it is straightforward to verify that truth-telling is always optimal for player i, and her ex-post net utility from this strategy is always non-negative. ■

θ'_{-i}	θ'_i	net utility of θ_i
$\theta'_{-i} = \theta_{-i} \geqq k$	$\theta'_i \geqq k$	0
	$\theta'_i < k$	$r_i(k - \theta_i)$
$\theta'_{-i} = \theta_{-i} < k$	$\theta'_i \geqq \theta_{-i}$	$r_{-i}(\theta_i - \theta_{-i}) + r_i(k - \theta_{-i})$
	$\theta'_i < \theta_{-i}$	$r_i(k - \theta_i)$

Table B.1: Net utility of type θ_i conditional on reports θ'_i and $\theta'_{-i} = \theta_{-i}$

B.3 Proof of Lemma 2.2

The result in Lemma 2.2 follows from arguments similar to those in Williams (1999) and Fieseler et al. (2003). In particular, revenue equivalence implies that any two efficient and IC mechanisms induce the same interim expected transfer, up to a constant. For any dispute (r_1, F, k), therefore, an efficient, IR, IC and BB mechanism exists if and only if for any efficient IC mechanism Γ it holds that

$$U_1^\Gamma(\theta_1^{*\Gamma}(r_1,k); r_1, k) + U_2^\Gamma(\theta_2^{*\Gamma}(r_2,k); r_2, k) \geqq \mathbb{E}_\theta(t_1^\Gamma(\theta_1,\theta_2) + t_2^\Gamma(\theta_1,\theta_2)),$$

where $\theta_i^{*\Gamma}, U_i^\Gamma, t_i^\Gamma$ denote i's worst-off type, net expected utility and payments under Γ. From Lemma 2.1, Γ^* is efficient and IC, which yields the necessary and sufficient condition $\mathcal{L}(r_1, r_2, k) \geqq \mathcal{S}(k)$.

Finally, to see that $\theta_i^*(r_i, k) = F^{-1}(r_i) \wedge k$, first note that

$$U_i^\Gamma(\theta_i; r_i, k) = \begin{cases} \displaystyle\int_0^{\theta_i} \theta_i - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) - r_i(\theta_i - k) & \text{if } \theta_i < k, \\ \displaystyle\int_0^{k} \theta_i - \theta_{-i} \,\mathrm{d}F(\theta_{-i}) - F(k) r_i(\theta_i - k) & \text{if } \theta_i > k. \end{cases}$$

Therefore,

$$\frac{\partial U_i^\Gamma(\theta_i; r_i, k)}{\partial \theta_i} = \begin{cases} F(\theta_i) - r_i & \text{if } \theta_i < k, \\ (1 - r_i) F(k) & \text{if } \theta_i > k, \end{cases} \quad \text{and} \quad \frac{\partial^2 U_i^\Gamma(\theta_i; r_i, k)}{\partial \theta_i^2} = \begin{cases} f(\theta_i) & \text{if } \theta_i < k, \\ 0 & \text{if } \theta_i > k. \end{cases}$$

Note that U_i^Γ is strictly convex on $[0, k]$, strictly decreasing at $\theta_i = 0$, and linear and strictly increasing on $[k, 1]$. Therefore, using the first-order condition, it has a unique minimum at either $F^{-1}(r_i)$ or k. ■

B.4 Proof of Corollary 2.1

Part 1. First, using expression (2.3), the expected subsidy can be bounded from below as follows:

$$\begin{aligned} \mathcal{S}(k) &= F(k)\left(1 - k - \int_k^1 F(\theta)\,\mathrm{d}\theta\right) + \int_0^k F(\theta)\,\mathrm{d}\theta - \int_0^k F^2(\theta)\,\mathrm{d}\theta \\ &\geqq F(k)\left(1 - k - \int_k^1 F(\theta)\,\mathrm{d}\theta\right) + \int_0^k F(\theta)\,\mathrm{d}\theta - F(k)\int_0^k F(\theta)\,\mathrm{d}\theta \\ &= F(k)\left(1 - \int_0^1 F(\theta)\,\mathrm{d}\theta - k + \int_0^k \frac{F(\theta)}{F(k)}\,\mathrm{d}\theta\right) \\ &= F(k)\left(\int_0^1 1 - F(\theta)\,\mathrm{d}\theta - \int_0^k 1 - \frac{F(\theta)}{F(k)}\,\mathrm{d}\theta\right) \\ &= F(k)\left(\mathbb{E}\theta - \mathbb{E}(\theta | \theta < k)\right). \end{aligned} \tag{B.1}$$

Second, we can show that the lump-sum fee is maximized by the equal-share partnership. Using expression (2.4) and denoting the worst-off types in the equal-share partnership by $\theta^*(1/2,k) \doteq \theta_1^*(1/2,k) = \theta_2^*(1/2,k)$,

$$\begin{aligned}
\mathcal{L}\left(\frac{1}{2},\frac{1}{2},k\right) - \mathcal{L}(r_1,r_2,k) &= k - \int_0^{\theta^*(1/2,k)} 1/2 - F(\theta_2)\,\mathrm{d}\theta_2 - \int_0^{\theta^*(1/2,k)} 1/2 - F(\theta_1)\,\mathrm{d}\theta_1 \\
&\qquad - k + \int_0^{\theta_1^*(r_1,k)} r_1 - F(\theta_2)\,\mathrm{d}\theta_2 + \int_0^{\theta_2^*(r_2,k)} r_2 - F(\theta_1)\,\mathrm{d}\theta_1 \\
&= - \int_0^{\theta^*(1/2,k)} r_1 - F(\theta_2)\,\mathrm{d}\theta_2 - \int_0^{\theta^*(1/2,k)} r_2 - F(\theta_1)\,\mathrm{d}\theta_1 \\
&\qquad + \int_0^{\theta_1^*(r_1,k)} r_1 - F(\theta_2)\,\mathrm{d}\theta_2 + \int_0^{\theta_2^*(r_2,k)} r_2 - F(\theta_1)\,\mathrm{d}\theta_1 \\
&= \int_{\theta^*(1/2,k)}^{\theta_1^*(r_1,k)} r_1 - F(\theta_2)\,\mathrm{d}\theta_2 - \int_{\theta_2^*(r_2,k)}^{\theta^*(1/2,k)} r_2 - F(\theta_1)\,\mathrm{d}\theta_1 \geqq 0,
\end{aligned}$$

In the second equality, we added $r_1 - 1/2$ to the first integrand and subtracted $r_1 - 1/2$ from the second integrand. To show the last inequality, assume withouth loss of generality that $r_1 \geqq 1/2 \geqq r_2$. If $\theta_2^*(r_2,k) = k$ holds, then $\theta_1^*(r_1,k) = \theta^*(1/2,k) = k$, and both integrals are 0. If $\theta_2^*(r_2,k) = F^{-1}(r_2) < k$, the last inequality follows from $r_1 \geqq F(\theta)$ for $\theta \leqq \theta_1^*(r_1,k) \leqq F^{-1}(r_1)$, and $r_2 \leqq F(\theta)$ for $\theta \geqq \theta_2^*(r_2,k) = F^{-1}(r_2)$.

Using this observation, the lump-sum fee for an arbitrary share allocation can be bounded from above:

$$
\begin{aligned}
\mathcal{L}(r_1, r_2, k) \leqq \mathcal{L}(1/2, 1/2, k) &= k - 2\int_0^{k \wedge \mathrm{Med}(\theta)} 1/2 - F(\theta)\,\mathrm{d}\theta \\
&= \begin{cases} k - 2F(k)\displaystyle\int_0^k \frac{1/2}{F(k)} - 1 + 1 - \frac{F(\theta)}{F(k)}\,\mathrm{d}\theta & \text{if } k \leqq \mathrm{Med}(\theta), \\ k - 2F(\mathrm{Med}(\theta))\displaystyle\int_0^{\mathrm{Med}(\theta)} 1 - \frac{F(\theta)}{F(\mathrm{Med}(\theta))}\,\mathrm{d}\theta & \text{if } k > \mathrm{Med}(\theta), \end{cases} \\
&= \begin{cases} 2F(k)\left(k - \mathbb{E}\Big(\theta|\theta < k\Big)\right) & \text{if } k \leqq \mathrm{Med}(\theta), \\ k - \mathbb{E}(\theta|\theta < \mathrm{Med}(\theta)) & \text{if } k > \mathrm{Med}(\theta), \end{cases} \\
&\leqq 2F(k)\left(k - \mathbb{E}\Big(\theta|\theta < k\Big)\right).
\end{aligned} \tag{B.2}
$$

The last inequality is trivially satisfied (as an equality) in the $k \leqq \mathrm{Med}(\theta)$ case. For $k > \mathrm{Med}(\theta)$, we can use the law of iterated expectations for $\mathbb{E}\Big(\theta|\theta < k\Big)$ and the partition $[0, k) = [0, \mathrm{Med}(\theta)) \cup [\mathrm{Med}(\theta), k)$:

$$
\begin{aligned}
&2F(k)\left(k - \mathbb{E}\Big(\theta|\theta < k\Big)\right) - \left(k - \mathbb{E}\Big(\theta|\theta < \mathrm{Med}(\theta)\Big)\right) \\
&= (2F(k) - 1)k - 2F(k)\left(\mathbb{E}\Big(\theta|\theta < k\Big) - \frac{1}{2F(k)}\mathbb{E}\Big(\theta|\theta < \mathrm{Med}(\theta)\Big)\right) \\
&= (2F(k) - 1)k - 2F(k)\Bigg(\frac{1/2}{F(k)}\mathbb{E}\Big(\theta|\theta < \mathrm{Med}(\theta)\Big) + \frac{F(k) - 1/2}{F(k)}\mathbb{E}\Big(\theta|\theta \in [\mathrm{Med}(\theta), k)\Big) \\
&\qquad\qquad\qquad\qquad - \frac{1}{2F(k)}\mathbb{E}\Big(\theta|\theta < \mathrm{Med}(\theta)\Big)\Bigg) \\
&= (2F(k) - 1)\left(k - \mathbb{E}\Big(\theta|\theta \in [\mathrm{Med}(\theta), k)\Big)\right) > 0.
\end{aligned}
$$

Using the bounds (B.1) and (B.2) derived for $\mathcal{S}$ and $\mathcal{L}$ along with Lemma 2.2 establishes the sufficient condition stated in the corollary.

Part 2. The expected subsidy is bounded from above by

$$\mathcal{S}(k) = F(k)\int_k^1 1 - F(\theta)\,\mathrm{d}\theta + \int_0^k F(\theta)(1-F(\theta))\,\mathrm{d}\theta \leqq F(k)\int_0^1 1 - F(\theta)\,\mathrm{d}\theta = F(k)\mathbb{E}\theta,$$

where the last equality follows from integration by parts.

Similarly, for $r_1 = \frac{1}{2}$ and $k \geqq \mathrm{Med}(\theta)$,

$$\begin{aligned}\mathcal{L}\left(\frac{1}{2},\frac{1}{2},k\right) &= k - \mathrm{Med}(\theta) + 2\int_0^{\mathrm{Med}(\theta)} F(\theta)\,\mathrm{d}\theta = k - 2\int_0^{\mathrm{Med}(\theta)} \theta f(\theta)\,\mathrm{d}\theta \\ &= k - \mathbb{E}(\theta|\theta < \mathrm{Med}(\theta)),\end{aligned}$$

where the second equality again follows from integration by parts.

Therefore, by Lemma 2.2, the condition $k \geqq F(k)\mathbb{E}\theta + \mathbb{E}(\theta|\theta < \mathrm{Med}(\theta))$ is sufficient for the possibility of efficiently resolving the dispute. ■

B.5 Proof of Proposition 2.1

The derivative of the expected subsidy can be written as

$$\frac{\mathrm{d}\mathcal{S}(k)}{\mathrm{d}k} = f(k)\int_k^1 1 - F(\theta)\,\mathrm{d}\theta = f(k)\,(1-F(k))\,\mathbb{E}(\theta - k|\theta > k), \tag{B.3}$$

where the second equality follows from integration by parts. For sufficiently small k, it holds that $\theta_1^*(r_1,k) = \theta_2^*(r_2,k) = k$; hence, from (2.4), $\frac{\partial\mathcal{L}(r_1,r_2,k)}{\partial k} = 2F(k)$. Since f is bounded away from 0 and $F(k)$ vanishes as $k \to 0$, the budget surplus $\mathcal{L}(r_1,r_2,k) - \mathcal{S}(k)$ is decreasing on $(0,\underline{k})$, for some $\underline{k} \in (0,1)$. Similarly, for k large enough, $\theta_i^*(r_i,k) = F^{-1}(r_i)$, and from (2.4), $\frac{\partial\mathcal{L}(r_1,r_2,k)}{\partial k} = 1$. Since f is bounded, $\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}k}$ converges to 0 as $k \to 1$. Since $\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}k}$ vanishes but $\frac{\partial\mathcal{L}}{\partial k}$ does not as $k \to 1$, there exists $\overline{k} \in (\underline{k},1)$ such that the surplus is increasing on $(\overline{k},1)$. ■

B.6 Proof of Part 1 of Proposition 2.2

Assume without loss of generality that $r_1 \geqq 1/2 \geqq r_2$. Using formulas (2.3) and (2.4) for the expected subsidy and the lump-sum payment, it is easy to show that the difference in the budget surplus under G and F is equal to[1]

$$\begin{aligned}
\Delta(\mathcal{L}-\mathcal{S}) &= (F(k)-G(k))(1-k) + \int_0^k F(\theta) - G(\theta)\,\mathrm{d}\theta \\
&\quad - \left(F(k)\int_k^1 F(\theta)\,\mathrm{d}\theta - G(k)\int_k^1 G(\theta)\,\mathrm{d}\theta\right) - \int_0^k F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta \\
&\quad - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta - \int_{\theta_2^{*F}}^{\theta_2^{*G}} r_2 - G(\theta)\,\mathrm{d}\theta \\
&\quad - \int_0^{\theta_1^{*F}} F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^{\theta_2^{*F}} F(\theta) - G(\theta)\,\mathrm{d}\theta, \qquad \text{(B.4)}
\end{aligned}$$

where θ_i^{*F} and θ_i^{*G} denote the worst-off types of partner i under F and G.

Assume G first-order stochastically dominates F. Then $\theta_i^{*F} \leqq \theta_i^{*G}$, and $G(\theta) \leqq r_i$ for all $\theta \leqq \theta_i^{*G} = k \wedge G^{-1}(r_i)$. Therefore, the first two terms are non-negative while all other terms are non-positive in this expression. Using $F(k) = G(k)$ and dropping some of the non-positive terms we get

$$\begin{aligned}
\Delta(\mathcal{L}-\mathcal{S}) &\leqq \int_0^k F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^k F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta \\
&\quad - \int_0^{\theta_1^{*F}} F(\theta) - G(\theta)\,\mathrm{d}\theta \\
&= \int_{\theta_1^{*F}}^k F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^k F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta. \qquad \text{(B.5)}
\end{aligned}$$

[1] In this proof, we drop the arguments r_1, r_2 and k of the functions $\mathcal{L}$, $\mathcal{S}$, θ_i^{*F}, and θ_i^{*G} to ease the notation.

In the last line, the first term is again non-negative while all other two terms are non-positive. If $\theta_1^{*F} = k$, the first term disappears and we get the desired result. In the $\theta_1^{*F} < k$ case, the following algebra shows the nonnegativity of the difference in budget surplus:

$$
\begin{aligned}
\Delta(\mathcal{L} - \mathcal{S}) &\leqq \int_{\theta_1^{*F}}^{k} F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^{k} F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta \\
&= \int_{\theta_1^{*G}}^{k} (F(\theta) - G(\theta))(1 - (F(\theta) + G(\theta)))\mathrm{d}\theta \\
&\qquad + \int_{\theta_1^{*F}}^{\theta_1^{*G}} F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^{\theta_1^{*G}} F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta \\
&\leqq \int_{\theta_1^{*F}}^{\theta_1^{*G}} F(\theta) - G(\theta)\,\mathrm{d}\theta - \int_0^{\theta_1^{*G}} F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta - \int_{\theta_1^{*F}}^{\theta_1^{*G}} r_1 - G(\theta)\,\mathrm{d}\theta \\
&= \int_{\theta_1^{*F}}^{\theta_1^{*G}} F(\theta)(1 - F(\theta)) + G^2(\theta) - r_1\,\mathrm{d}\theta - \int_0^{\theta_1^{*F}} F^2(\theta) - G^2(\theta)\,\mathrm{d}\theta \\
&\leqq \int_{\theta_1^{*F}}^{\theta_1^{*G}} F(\theta)(1 - F(\theta)) + G^2(\theta) - r_1\,\mathrm{d}\theta \\
&\leqq \int_{\theta_1^{*F}}^{\theta_1^{*G}} F(\theta)(1 - F(\theta)) - r_1(1 - r_1)\,\mathrm{d}\theta \leqq 0.
\end{aligned}
\tag{B.6}
$$

The first inequality is simply a repetition of the previous formula.

The validity of the second inequality can be seen as follows. If $\theta_1^{*G} = k$, the first integral is zero. If $\theta_1^{*G} < k$, then $\theta_1^{*G} = G^{-1}(r_1)$ must hold. Therefore, $F(\theta) \geqq G(\theta) \geqq r_1 \geqq 1/2$ is true for every $\theta_1 \geqq \theta_1^{*G}$, guaranteeing the non-positivity of the integrand in the first integral. The third inequality drops a non-positive term, and the forth inequality follows from the fact that $G(\theta) \leqq r_1$ for $\theta \leqq \theta_1^{*G} \leqq G^{-1}(r_1)$. Finally, the last inequality holds since $F(\theta) \geqq r_1 \geqq 1/2$ for $\theta \geqq \theta_1^{*F} = F^{-1}(r_1)$ and the function $x \mapsto x(1-x)$ being decreasing for $x \geqq 1/2$ together guarantee the non-positivity of the integrand.

To show that the decrease in the budget surplus is strict, first notice that the continuity of the cdf's and stochastic dominance implies that the set $\{\theta \in [0,1] : F(\theta) > G(\theta)\}$ must be of positive measure. If F and G are different for some $\theta \in [k, 1]$, then the third term in (B.4) is negative, and dropping it makes the inequality in (B.5) strict. If F and G are different for some $\theta \in [0, k]$, then either $\theta_1^{*F} = k$, or $\theta_1^{*F} < \theta_1^{*G} \leqq k$, or $\theta_1^{*F} = \theta_1^{*G} < k$. In the first case, the right-hand side of the first inequality in (B.6) is already negative since the first and third terms are 0, and the second term is negative. In the second case, the last inequality in (B.6) must be strict since the integrand is negative for $\theta \in (\theta_1^{*F}, \theta_1^{*G}]$, and this set is of positive measure. Finally, consider the third case. If F and G are different on $[\theta_1^{*G}, k]$, then the second inequality in (B.6) is strict since the first integrand on the left-hand side is negative on a set of positive measure. Otherwise, the reasoning is the same as in the $\theta_1^{*F} = k$ case. ■

B.7 Proof of Proposition 2.4

To analyze the effect of a small change in the threshold used for partner 1 on the budget balance, we compute the derivative of the new expected subsidy and the derivative of the new total lump-sum payment with respect to partner 1's threshold.

Formally, let $\mathcal{S}(l_1, l_2; k)$ and $\mathcal{L}(l_1, l_2; r_1, r_2, k)$ denote the expected subsidy and total lump-sum payment for partnership (r_1, F, k) when the thresholds l_1 and l_2 and the corresponding modified allocation rule and payment rules are used in the dissolution decision. To derive (2.14), we need to compute the partial derivatives $\frac{\partial}{\partial l_1}\mathcal{S}(l_1, l_2; k)$ and $\frac{\partial}{\partial l_1}\mathcal{L}(l_1, l_2; r_1, r_2, k)$ at $l_1 = l_2 = k$. A slight increase in the threshold used for player 1 weakly improves the budget balance if and only if the derivative of the lump-sum payment is larger than that of the expected subsidy.

The new allocation rule and the implementing payment rules, and consequently the new expected subsidy and the total lump-sum fee functions (and their derivatives) are slightly different for the $l_1 \geqq k$ and $l_1 \leqq k$ cases. Here, we derive these functions only for the $l_1 \geqq k$ case to compute the right derivatives and later argue that the same expression can be derived for $l_1 \leqq k$.

New allocation and payment rule Assume partner 1's threshold is raised from k to some $l_1 > k$. The new allocation rule (illustrated in Figure 2.5 (a)) is defined as follows:

$$q^{l_1}(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 \geqq l_1, \theta_2 \geqq k, \\ d_2 & \text{if } \theta_1 < l_1, \theta_2 > \theta_1, \\ d_1 & \text{otherwise.} \end{cases}$$

It is straightforward to check that the following transfer rules (illustrated in Figures 2.5 (b) and 2.5 (c)) implement this allocation rule in ex-post equilibrium:

$$t_1^{l_1}(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 \geqq l_1, \theta_2 \geqq k, \\ r_1\theta_2 + r_1(l_1 - k) & \text{if } \theta_1 < l_1, \theta_2 > l_1, \\ r_1\theta_2 + r_1(\theta_2 - k) + r_2(\theta_2 - l_1) & \text{if } \theta_1 < \theta_2, \theta_2 \in (k, l_1), \\ r_1\theta_2 & \text{if } \theta_1 < \theta_2, \theta_2 < k, \\ -r_2 l_1 + r_1(\theta_2 - k) & \text{if } \theta_1 \in (\theta_2, l_1), \theta_2 \in (k, l_1), \\ -r_2\theta_2 & \text{if } \theta_1 > \theta_2, \theta_2 < k, \end{cases}$$

$$t_2^{l_1}(\theta_1,\theta_2) = \begin{cases} 0 & \text{if } \theta_1 \geqq l_1, \theta_2 \geqq k, \\ -r_1\theta_1 & \text{if } \theta_1 < l_1, \theta_2 > l_1, \\ -r_1\theta_1 & \text{if } \theta_1 < \theta_2, \theta_2 \in (k, l_1), \\ -r_1\theta_1 & \text{if } \theta_1 < \theta_2, \theta_2 < k, \\ r_2\theta_1 & \text{if } \theta_1 \in (\theta_2, l_1), \theta_2 \in (k, l_1), \\ r_2\theta_1 & \text{if } \theta_1 > \theta_2, \theta_2 < k. \end{cases}$$

Expected subsidy The partial derivative of the difference between the expected subsidy for the $l_1 > k$ and the $l_1 = k$ cases with respect to l_1 is the same as that of the expected subsidy for $l_1 > k$. It is useful to consider the former expression to simplify the derivation. The difference in the expected subsidy is

$$\begin{aligned} \Delta\mathcal{S}(l_1,k;k) &\doteq \mathcal{S}(l_1,k;k) - \mathcal{S}(k) \\ &= \int_0^k \int_{l_1}^1 r_1(l_1-k)\,\mathrm{d}F(\theta_2)\,\mathrm{d}F(\theta_1) \\ &+ \int_0^k \int_k^{l_1} r_1(\theta_2-k) + r_2(\theta_2-l_1)\,\mathrm{d}F(\theta_2)\,\mathrm{d}F(\theta_1) \\ &+ \int_k^{l_1} \int_{l_1}^1 r_1(\theta_2-\theta_1) + r_1(l_1-k)\,\mathrm{d}F(\theta_2)\,\mathrm{d}F(\theta_1) \\ &+ \int_k^{l_1} \int_{\theta_1}^{l_1} r_1(\theta_2-\theta_1) + r_1(\theta_2-k) + r_2(\theta_2-l_1)\,\mathrm{d}F(\theta_2)\,\mathrm{d}F(\theta_1) \\ &+ \int_k^{l_1} \int_k^{\theta_1} r_2(\theta_1-l_1) + r_1(\theta_2-k)\,\mathrm{d}F(\theta_2)\,\mathrm{d}F(\theta_1). \end{aligned}$$

Using the Leibniz integral rule, we can compute the (right) derivative of this function with respect to l_1 at $l_1 = k$:

$$\frac{\partial}{\partial l_1}\Delta\mathcal{S}(l_1,k;k)\bigg|_{l_1=k} = r_1\Big(F(k)(1-F(k)) + f(k)\mathbb{E}((\theta-k)\vee 0)\Big). \tag{B.7}$$

Net expected utility of partner 2 To analyze the change in the total lump-sum payment, we need to the derive the net expected utility function of both partners when an $l_1 > k$ threshold is used for partner 1 in the dissolution decision. The net expected utility curve of partner 2 is:[2]

$$U_2(\theta_2;r_2,l_1,k,k) =$$
$$\begin{cases} \displaystyle\int_0^{\theta_2} \theta_2 - r_1\theta_1 \,\mathrm{d}F(\theta_1) + \int_{\theta_2}^{1} r_2\theta_1 \,\mathrm{d}F(\theta_1) - \int_0^{1} r_2(\theta_1+\theta_2-k)\,\mathrm{d}F(\theta_1) & \text{if } \theta_2 \leqq k, \\ \displaystyle\int_0^{\theta_2} \theta_2 - r_1\theta_1 \,\mathrm{d}F(\theta_1) + \int_{\theta_2}^{l_1} r_2\theta_1 \,\mathrm{d}F(\theta_1) - \int_0^{l_1} r_2(\theta_1+\theta_2-k)\,\mathrm{d}F(\theta_1) & \text{if } \theta_2 \in [k,l_1], \\ \displaystyle\int_0^{l_1} \theta_2 - r_1\theta_1 \,\mathrm{d}F(\theta_1) - \int_0^{l_1} r_2(\theta_1+\theta_2-k)\,\mathrm{d}F(\theta_1) & \text{if } \theta_2 \geqq l_1. \end{cases}$$

The partial derivative of this function with respect to θ_2:

$$\frac{\partial U_2}{\partial \theta_2}(\theta_2;r_1,r_2,l_1,k,k) = \begin{cases} F(\theta_2) - r_2 & \text{if } \theta_2 < k, \\ F(\theta_2) - r_2F(l_1) & \text{if } \theta_2 \in (k,l_1), \\ r_1F(l_1) & \text{if } \theta_2 > l_1. \end{cases}$$

The partial derivative is always positive for $\theta_2 > l_1$, and also for $\theta_2 \in (k,l_1)$ as long as l_1 is sufficiently close to k. Moreover, increasing the threshold used for partner 1 does not change partner 2's net expected utility curve for $\theta_2 \leqq k$. These observations imply that, when l_1 is

[2]The superscript of U_i indicating the efficient allocation is omitted throughout the proof to simplify the notation.

sufficiently close to k, partner 2's worst-off type and the worst-off type's net expected utility are the same as in the $l_1 = k$ case. In other words, partner 2 is willing to pay the same amount to participate. As $l_1 \to k$, this implies that

$$\frac{\mathrm{d}U_2}{\mathrm{d}l_1}(\theta_2^*(r_1, r_2; l_1, k, k); r_1, r_2, l_1, k, k)\Big|_{l_1=k} = 0. \tag{B.8}$$

Net expected utility of partner 1 Partner 1's net expected utility when threshold $l_1 > k$ is used in the dissolution decision:

$$U_1(\theta_2; r_1, r_2, l_1, k, k) =$$
$$\begin{cases} \displaystyle\int_0^{\theta_1} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) + \int_{\theta_1}^{k} r_1\theta_2 \,\mathrm{d}F(\theta_2) \\ \displaystyle\qquad + \int_k^{l_1} r_1\theta_2 + r_1(\theta_2 - k) + r_2(\theta_2 - l_1)\,\mathrm{d}F(\theta_2) \\ \displaystyle\qquad + \int_{l_1}^{1} r_1\theta_2 + r_1(l_1 - k)\,\mathrm{d}F(\theta_2) - \int_0^1 r_1(\theta_1 + \theta_2 - k)\,\mathrm{d}F(\theta_2) & \text{if } \theta_1 \leqq k, \\ \displaystyle\int_0^{k} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) + \int_k^{\theta_1} \theta_1 - r_2 l_1 + r_1(\theta_2 - k)\,\mathrm{d}F(\theta_2) \\ \displaystyle\qquad + \int_{\theta_1}^{l_1} r_1\theta_2 + r_1(\theta_2 - k) + r_2(\theta_2 - l_1)\,\mathrm{d}F(\theta_2) \\ \displaystyle\qquad + \int_{l_1}^{1} r_1\theta_2 + r_1(l_1 - k)\,\mathrm{d}F(\theta_2) - \int_0^1 r_1(\theta_1 + \theta_2 - k)\,\mathrm{d}F(\theta_2) & \text{if } \theta_1 \in [k, l_1], \\ \displaystyle\int_0^{k} \theta_1 - r_2\theta_2 \,\mathrm{d}F(\theta_2) - \int_0^k r_1(\theta_1 + \theta_2 - k)\,\mathrm{d}F(\theta_2) & \text{if } \theta_1 \geqq l_1. \end{cases}$$

The derivative of the net expected utility with respect to θ_1:

$$\frac{\partial U_1}{\partial \theta_1}(\theta_1; r_1, r_2, l_1, k, k) = \begin{cases} F(\theta_1) - r_1 & \text{if } \theta_2 < k, \\ F(\theta_1) - r_1 & \text{if } \theta_2 \in (k, l_1), \\ r_2 F(k) & \text{if } \theta_2 > l_1. \end{cases}$$

Since $r_2F(k) > 0$, the net expected utility will be minimized at some $\theta_2 \leqq l_1$. If $F(k) \geqq r_1$, the worst-off type is equal to $F^{-1}(r_1)$ and belongs to $[0, k]$. Since the net expected utility function is differentiable and the first-order condition is satisfied at this point, we can use the envelope theorem to compute the derivative of the worst-off type's net expected utility with respect to l_1. If $F(k) < r_1$ and l_1 is close enough to k, the net expected utility function is decreasing on $[0, l_1]$, and therefore the worst-off type is equal to l_1. However, the net expected utility function is not differentiable at this point. Instead of directly applying the envelope theorem, we can plug in the worst-off type as a function of l_1 into the net expected utility and take the total derivative with respect to the parameter l_1. The resulting expression is:

$$\frac{\mathrm{d}U_1}{\mathrm{d}l_1}(\theta_1^*(r_1, r_2; l_1, k, k); r_1, r_2, l_1, k, k) = \begin{cases} r_1(1 - F(l_1)) - r_2(F(l_1) - F(k)) & \text{if } r_1 \leqq F(k), \\ r_2F(k) & \text{if } r_1 > F(k). \end{cases}$$

As $l_1 \to k$, this expression becomes

$$\frac{\mathrm{d}U_1}{\mathrm{d}l_1}(\theta_1^*(r_1, r_2; l_1, k, k); r_1, r_2, l_1, k, k)\Big|_{l_1=k} = \begin{cases} r_1(1 - F(k)) & \text{if } r_1 \leqq F(k), \\ r_2F(k) & \text{if } r_1 > F(k). \end{cases} \tag{B.9}$$

Lump-sum payment Adding up expressions (B.8) and (B.9), we get the derivative of the total lump-sum fee with respect to l_1 at $l_1 = k$:

$$\frac{\partial}{\partial l_1}\mathcal{L}(l_1, k; r_1, r_2, k)\Big|_{l_1=k} = \begin{cases} r_1(1 - F(k)) & \text{if } r_1 \leqq F(k), \\ r_2F(k) & \text{if } r_1 > F(k). \end{cases} \tag{B.10}$$

Comparison of the right derivatives A slight increase in l_1 from $l_1 = k$ increases the budget surplus if and only if the lump-sum fee increases more than the expected subsidy.

Comparing (B.7) and (B.10), this translates to the following condition:

$$r_1\Big(F(k)(1-F(k)) + f(k)\mathbb{E}((\theta-k)\vee 0)\Big) \gtreqqless \begin{cases} r_1(1-F(k)) & \text{if } r_1 \leqq F(k), \\ r_2F(k) & \text{if } r_1 > F(k). \end{cases} \tag{B.11}$$

The same steps can be used to derive the analogous condition for partner 2.

Left derivatives and the second part of the statement Although the allocation rule and the payment rules are different in the $l_1 \leqq k$ case, it is easy to check that this difference is of second order and consequently disappears from the derivatives as l_1 converges to k. Therefore, condition (B.11) also holds for the left derivatives. The same observation is true if partner 2's threshold changes.

The validity of condition (2.14) does not imply immediately that the same condition can be used to evaluate the effects of simultaneous changes in the two thresholds. This condition is based on the partial derivatives $\frac{\partial}{\partial l_i}\mathcal{S}(l_1, l_2; k)$ and $\frac{\partial}{\partial l_i}\mathcal{L}(l_1, l_2; r_1, r_2, k)$ at $l_1 = l_2 = k$. To use the same condition for simultaneous changes, we need to show the differentiability of $\mathcal{S}(l_1, l_2; k)$ and $\mathcal{L}(l_1, l_2; r_1, r_2, k)$ as multivariate functions of (l_1, l_2) at (k, k). The continuity of these partial derivatives in l_1 and l_2 in a neighborhood of (k, k) guarantees multivariate differentiability. It is easy to verify that if $k \neq F^{-1}(r_1)$, $F^{-1}(r_2)$, continuity is satisfied for subsets of a small neighborhood in which the order of l_1, l_2, and k is fixed. Therefore, it remains only to check the boundaries between these sets. Indeed, we can make the same observation as above: the difference in the expected subsidy or the lump-sum fee function between any pair of neighboring regions is of second order, and will disappear as we approach the boundary separating the two regions. This guarantees the multivariate differentiability of the two functions and proves the second part of the proposition. ■

B.8 General partnership functions

This section contains the analysis for Section 2.5, which studies general partnership functions.

First note that given the properties of V, the threshold functions are increasing, and twice differentiable almost everywhere. Furthermore, $\underline{\theta}_i$ is convex and $\overline{\theta}_i$ concave whenever they are not equal to 0, 1 and do not intersect with the 45-degree line.

For each $i = 1, 2$, denote partner i's net expected utility by $U_i(\theta_i; r_i, V)$, where

$$U_i(\theta_i; r_i, V) = \int_0^{\underline{\theta}_{-i}(\theta_i)} \theta_i - \overline{\theta}_i(\theta_{-i}) - r_i\Big(V(\theta_1, \theta_2) - \overline{\theta}_i(\theta_{-i})\Big)\, \mathrm{d}F_{-i}(\theta_{-i}) + \\ + \int_{\overline{\theta}_{-i}(\theta_i)}^1 r_i\left(\theta_{-i} - V(\theta_1, \theta_2)\right) \mathrm{d}F_{-i}(\theta_{-i}). \tag{B.12}$$

Given the mechanism Γ^* defined by the allocation rule (2.16) and the payment rule (2.17), denote the expected subsidy the arbitrator must incur by

$$\mathcal{S}(r_1, r_2, V) = r_1 \int_0^1 \int_{\overline{\theta}_2(\theta_1)}^1 \theta_2 - \overline{\theta}_2(\theta_1)\, \mathrm{d}F_2(\theta_2)\, \mathrm{d}F_1(\theta_1) \\ + r_2 \int_0^1 \int_0^{\underline{\theta}_2(\theta_1)} \theta_1 - \overline{\theta}_1(\theta_2)\, \mathrm{d}F_2(\theta_2)\, \mathrm{d}F_1(\theta_1). \tag{B.13}$$

Define the worst-off type of each agent i as $\theta_i^*(r_i, V) \in \mathrm{argmin}_{\theta_i \in [0,1]} U_i(\theta_i; r_i, V)$, and let $\mathcal{L}(r_1, r_2, V) \doteq U_1(\theta_1^*(r_1, V); r_1, V) + U_2(\theta_2^*(r_2, V); r_2, V)$ denote the largest lump-sum fee that can be charged from the agents without violating their participation constraints (i.e., the sum of the maximal participation fees the agents are willing to pay).

The following lemma summarizes several useful properties of the net expected utility and the worst-off type.

Lemma B.1 Given the mechanism Γ^*, the following properties hold:

1. $\frac{\partial U_i(\theta_i; r_i, V)}{\partial \theta_i}$ is non-increasing in r_i;
2. For all $r_i \in [0,1]$, the function $U_i(\theta_i; r_i, V)$ is convex in θ_i;
3. For all $r_i \in [0,1]$, the function $\theta_i^*(r_i, V)$ is non-decreasing in r_i;
4. $U_i(\theta_i^*(r_i, V); r_i, V)$ is a concave function of r_i.

PROOF. Using the Leibniz rule to compute the first derivative of $U_i(\theta_i; r_i, V)$ with respect to θ_i, we obtain

$$\frac{\partial U_i(\theta_i; r_i, V)}{\partial \theta_i} = \int_0^{\underline{\theta}_{-i}(\theta_i)} 1 - r_i \frac{\partial V(\theta_1, \theta_2)}{\partial \theta_i} \, \mathrm{d}F_{-i}(\theta_{-i}) - \int_{\overline{\theta}_{-i}(\theta_i)}^1 r_i \frac{\partial V(\theta_1, \theta_2)}{\partial \theta_i} \, \mathrm{d}F_{-i}(\theta_{-i}).$$

Differentiating with respect to r_i, part 1 is immediate. For part 2, computing the second derivative gives

$$\begin{aligned} \frac{\partial^2 U_i(\theta_i; r_i, V)}{\partial \theta_i^2} &= \\ &- \int_0^{\underline{\theta}_{-i}(\theta_i)} r_i \frac{\partial^2 V(\theta_1, \theta_2)}{\partial \theta_i^2} \, \mathrm{d}F_{-i}(\theta_{-i}) + \left(1 - r_i \frac{\partial V(\theta_i, \underline{\theta}_{-i}(\theta_i))}{\partial \theta_i}\right) f_{-i}(\underline{\theta}_{-i}(\theta_i)) \frac{\mathrm{d}\underline{\theta}_{-i}(\theta_i)}{\mathrm{d}\theta_i} \\ &- \int_{\overline{\theta}_{-i}(\theta_i)}^1 r_i \frac{\partial^2 V(\theta_1, \theta_2)}{\partial \theta_i^2} \, \mathrm{d}F_{-i}(\theta_{-i}) + r_i \frac{\partial V(\theta_i, \overline{\theta}_{-i}(\theta_i))}{\partial \theta_i} f_{-i}(\overline{\theta}_{-i}(\theta_i)) \frac{\mathrm{d}\overline{\theta}_{-i}(\theta_i)}{\mathrm{d}\theta_i}. \end{aligned}$$

The concavity of V, the assumption that $\frac{\partial V(\theta_1, \theta_2)}{\partial \theta_i} \in (0,1)$ for all θ_i and θ_{-i}, and the observation that $\frac{\mathrm{d}\overline{\theta}_{-i}(\theta_i)}{\mathrm{d}\theta_i}, \frac{\mathrm{d}\underline{\theta}_{-i}(\theta_i)}{\mathrm{d}\theta_i} \geqq 0$ for almost every θ_i guarantee that this expression is non-negative, proving part 2. Part 3 is now straightforward from parts 1 and 2.

For part 4, let $r_1, r_1' \in [0,1]$ and $\lambda \in [0,1]$. Using the linearity of the net expected utility in r and the definition of θ^*,

$$\begin{aligned}
&U_i\Big(\theta_i^*(\lambda r_i + (1-\lambda) r_i', V); \lambda r_i + (1-\lambda) r_i', V\Big) \\
&= \lambda U_i\Big(\theta_i^*(\lambda r_i + (1-\lambda) r_i', V); r_i, V\Big) + (1-\lambda) U_i\Big(\theta_i^*(\lambda r_i + (1-\lambda) r_i', V); r_i', V)\Big) \\
&\geqq \lambda U_i(\theta_i^*(r_i, V); r_i, V) + (1-\lambda) U_i(\theta_i^*(r_i', V); r_i', V).
\end{aligned}$$

■

As in the analysis in Section 2.2, whether or not a dispute can be resolved efficiently hinges on the relationship between the net expected utility of the worst-off types and the expected subsidy under Γ^*.

Lemma B.2 The partnership dispute (r_1, F_1, F_2, V) can be resolved efficiently if and only if $\mathcal{L}(r_1, r_2, V) \geqq \mathcal{S}(r_1, r_2, V)$.

The proof follows the same arguments as the one for Lemma 2.2. It is useful to rearrange the net expected utility as follows

$$\begin{aligned}
U_i(\theta_i; r_i, V) &= r_i \int_0^1 \Big(\max\{\theta_i, \theta_{-i}, V(\theta_1, \theta_2)\} - V(\theta_1, \theta_2)\Big) \mathrm{d}F_{-i}(\theta_{-i}) \\
&\quad + r_{-i} \int_0^1 \max\Big\{0, \theta_i - \overline{\theta}_i(\theta_{-i})\Big\} \mathrm{d}F_{-i}(\theta_{-i}) \\
&= r_i \mathbb{E}_{\theta_{-i}} \operatorname{Surp}(\theta_1, \theta_2; V) + r_{-i} \mathbb{E}_{\theta_i} \operatorname{Impr_i}(\theta_1, \theta_2, V), \qquad \text{(B.14)}
\end{aligned}$$

where $\operatorname{Surp}(\theta_1, \theta_2; V) \doteq \max\{\theta_1, \theta_2, V(\theta_1, \theta_2)\} - V(\theta_1, \theta_2)$ is the ex-post surplus from resolving the partnership dispute, and $\operatorname{Impr_i}(\theta_1, \theta_2, V) \doteq \max\Big\{0, \theta_i - \overline{\theta}_i(\theta_{-i})\Big\}$ is equal to the ex-post improvement generated by partner i's sole ownership of the asset relative to the value of the best alternative (the best effective partnership or $-i$'s sole ownership), when this improvement is positive, and 0 otherwise.

We can now study how the severity of a dispute, as defined formally below, affects the possibility of its efficient resolution.

Definition B.1 We say that the dispute in partnership W is less severe than the dispute in partnership V if $W(\theta_1, \theta_2) \geqq V(\theta_1, \theta_2)$ for all $\theta_1, \theta_2 \in [0, 1]$.

Lemma B.3 Assume that the dispute in partnership W is less severe than the dispute in partnership V.

1. The region of efficient dissolution for partnership W is a subset of the efficient region for partnership V. Denoting the analogous thresholds for W by $\underline{\omega}_i$ and $\overline{\omega}_i$, it holds that $\underline{\omega}_i \leqq \underline{\theta}_i$ and $\overline{\theta}_i \leqq \overline{\omega}_i$, for each $i = 1, 2$.

2. For all θ_i and i, the net expected utility is greater under V: $U_i(\theta_i; r_i, V) \geqq U_i(\theta_i; r_i, W)$. Consequently, the largest lump-sum fee that can be charged is smaller for a less severe dispute.

3. The expected subsidy is greater under V: $\mathcal{S}(r_1, r_2, V) \geqq \mathcal{S}(r_1, r_2, W)$.

Proof. Part 1 is true by the definition of the threshold functions. Part 2 immediately follows from the first point and equation (B.14), since both integrands are lower for partnership W. Part 3 is a consequence of equation (B.13); smaller functions are integrated over smaller sets in the case of partnership W. ■

The severity of a dispute affects both the value of the partnership when kept in tact and the region of dissolution. As the following proposition shows, efficient resolution crucially depends on the interactions between these two effects.

Proposition B.1 Assume that the dispute in partnership W is less severe than in partnership V.

1. If the region of efficient dissolution is the same for partnerships V and W (i.e., $\overline{\theta} = \overline{\omega}$ and $\underline{\theta} = \underline{\omega}$), then a partnership dispute is less costly to resolve for partnership V.

2. Otherwise, a less severe dispute might be less costly to resolve.

Proof.

Part 1. Note first that the expected subsidy requires information only on the threshold functions. Since the dissolution thresholds are identical for the two partnerships, the expected subsidy paid is the same for V and W. The lump-sum fee, however, is weakly larger for the more severe dispute according to part 2 of Lemma B.3. Therefore, the partnership dispute under V is less costly to resolve than the partnership dispute under W.

Part 2. Consider the extreme case in which partner 2 owns the partnership completely (i.e., $r_2 = 1$). Rewriting equation (B.12) for this case gives:

$$U_1(\theta_1; 0, V) = \int_0^{\underline{\theta}_2(\theta_1)} \theta_1 - \overline{\theta}_1(\theta_2)\,\mathrm{d}F_2(\theta_2),$$
$$U_2(\theta_2; 1, V) = \int_0^{\underline{\theta}_1(\theta_2)} \theta_2 - V(\theta_1, \theta_2)\,\mathrm{d}F_1(\theta_1) + \int_{\overline{\theta}_1(\theta_2)}^{1} \theta_1 - V(\theta_1, \theta_2)\,\mathrm{d}F_1(\theta_1).$$

Note that $\theta_1 = 0$ is definitely a worst-off type of player 1. First, the net expected utility of player 1 is non-negative by the individual rationality of the implementing mechanism. Second, by assumption, $0 \leqq \underline{\theta}_2(\theta_1) \leqq \theta_1$; therefore, $\underline{\theta}_2(0) = 0$ must be true. Consequently, $U_1(0; 0, V) = 0$; a partner without a share is never willing to pay any positive lump-sum fee ex ante.

Fix two value functions ($1 > \varepsilon > 0$):

$$\begin{aligned} W(\theta_1, \theta_2) &\doteq \min\left\{(1-\varepsilon)\theta_1 + \frac{\varepsilon}{2}(1+\theta_2), \frac{1+\theta_2}{2}\right\}, \\ V(\theta_1, \theta_2) &\doteq \min\left\{(1-\varepsilon)\theta_1 + \frac{\varepsilon}{2}(1+\theta_2), \frac{1+\theta_2}{2}\right\} - \frac{\varepsilon}{4}(1-\theta_2). \end{aligned}$$

First, note that both V and W fit our framework: they are both piecewise linear, concave functions, increasing in both types. By definition, the dispute in W less severe than in V for any $\varepsilon > 0$. For a given type θ_2, the threshold values are

$$\begin{aligned} \underline{\omega}_1(\theta_2) &= \max\left\{\frac{1-\varepsilon/2}{1-\varepsilon}\theta_2 - \frac{\varepsilon/2}{1-\varepsilon}, 0\right\}, \\ \underline{\theta}_1(\theta_2) &= \max\left\{\frac{1-3\varepsilon/4}{1-\varepsilon}\theta_2 - \frac{\varepsilon/4}{1-\varepsilon}, 0\right\}, \\ \overline{\omega}_1(\theta_2) &= \frac{1+\theta_2}{2}, \\ \overline{\theta}_1(\theta_2) &= \frac{1+3\theta_2}{4}. \end{aligned}$$

The threshold functions are illustrated in Figure B.1. The two shaded regions represent the two different linear segments in the definitions of the functions V and W.

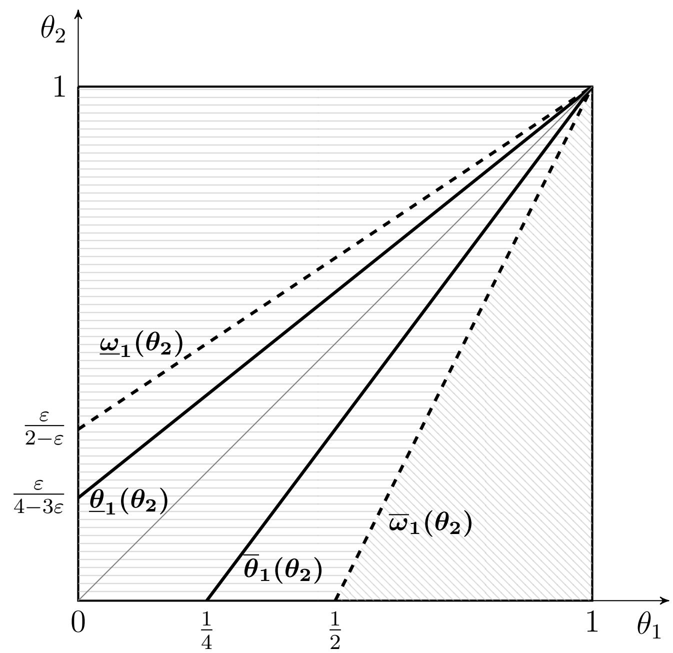


Figure B.1: Threshold functions used in the proof of Proposition B.1

Compute the difference in net expected utility for partner 2 under V and W:

$$
\begin{aligned}
&U_2(\theta_2;1,V) - U_2(\theta_2;1,W)\\
&\quad= \int_0^{\underline{\omega}_1(\theta_2)} W(\theta_1,\theta_2) - V(\theta_1,\theta_2)\,\mathrm{d}F_1(\theta_1) + \int_{\underline{\omega}_1(\theta_2)}^{\underline{\theta}_1(\theta_2)} \theta_2 - V(\theta_1,\theta_2)\,\mathrm{d}F_1(\theta_1)\\
&\qquad + \int_{\bar{\theta}_1(\theta_2)}^{\overline{\omega}_1(\theta_2)} \theta_1 - V(\theta_1,\theta_2)\,\mathrm{d}F_1(\theta_1) + \int_{\overline{\omega}_1(\theta_2)}^{1} W(\theta_1,\theta_2) - V(\theta_1,\theta_2)\,\mathrm{d}F_1(\theta_1)\\
&\quad= \int_0^{\underline{\omega}_1(\theta_2)} \underbrace{\frac{\varepsilon}{4}(1-\theta_2)}_{\leqq\frac{\varepsilon}{4}}\,\mathrm{d}F_1(\theta_1) + \int_{\underline{\omega}_1(\theta_2)}^{\underline{\theta}_1(\theta_2)} \underbrace{-\frac{\varepsilon}{4} - (1-\varepsilon)\theta_1 + \left(1-\frac{3\varepsilon}{4}\right)\theta_2}_{\leqq\frac{\varepsilon}{4}(1-\theta_2)\leqq\frac{\varepsilon}{4},\text{ using }\theta_1\geqq\underline{\omega}_1(\theta_2)}\,\mathrm{d}F_1(\theta_1)\\
&\qquad + \int_{\bar{\theta}_1(\theta_2)}^{\overline{\omega}_1(\theta_2)} \underbrace{-\frac{\varepsilon}{4} + \varepsilon\theta_1 - \frac{3\varepsilon}{4}\theta_2}_{\leqq\frac{\varepsilon}{4}(1-\theta_2)\leqq\frac{\varepsilon}{4},\text{ using }\theta_1\leqq\overline{\omega}_1(\theta_2)}\,\mathrm{d}F_1(\theta_1) + \int_{\overline{\omega}_1(\theta_2)}^{1} \underbrace{\frac{\varepsilon}{4}(1-\theta_2)}_{\leqq\frac{\varepsilon}{4}}\,\mathrm{d}F_1(\theta_1)\\
&\quad\leqq \frac{\varepsilon}{4}\left(\int_0^{\underline{\theta}_1(\theta_2)} 1\,\mathrm{d}F_1(\theta_1) + \int_{\bar{\theta}_1(\theta_2)}^{1} 1\,\mathrm{d}F_1(\theta_1)\right) \leqq \frac{\varepsilon}{4}.
\end{aligned}
$$

Hence, the difference in the net expected utility for *all types* of partner 2 can be made arbitrarily small by picking a small enough ε. Therefore, the change in the lump-sum fee partner 2 is willing to make can never be more than $\varepsilon/4$ either.

The difference in the expected subsidy can be computed using equation (B.13):

$$\begin{aligned}\mathcal{S}(r_1, r_2, V) - \mathcal{S}(r_1, r_2, W) = &\int_0^1 \int_0^{\underline{\theta}_2(\theta_1)} \theta_1 - \overline{\theta}_1(\theta_2)\,\mathrm{d}F_2(\theta_2)\,\mathrm{d}F_1(\theta_1) \\ &- \int_0^1 \int_0^{\underline{\omega}_2(\theta_1)} \theta_1 - \overline{\omega}_1(\theta_2)\,\mathrm{d}F_2(\theta_2)\,\mathrm{d}F_1(\theta_1).\end{aligned}$$

First, note that this formula involves only threshold functions that are below the diagonal in Figure 2.2. Therefore, this value does not depend on ε. Moreover, this difference is positive since $\overline{\omega}_1(\theta_2) > \overline{\theta}_1(\theta_2)$ for all $\theta_2 > 0$ and $\underline{\theta}_2(\theta_1) \geqq \underline{\omega}_2(\theta_1)$, and we also assume that both density functions are positive. Thus, there is a small enough ε such that the difference in the lump-sum fee collected is smaller than the difference in the expected subsidy paid. For such ε, a more severe partnership dispute is more costly to resolve. ■

Finally, the next result generalizes the findings of Proposition 2.1 by showing that partnership disputes that are not sufficiently severe cannot be resolved efficiently.

Proposition B.2 Fix a pair of threshold functions $\overline{\theta}_1$ and $\overline{\theta}_2$ that are not always equal to 0,

and a pair of type distributions F_1 and F_2. There exists a number $K > 0$ such that for every ownership structure (r_1, r_2), and for every partnership function V satisfying

(i) the dissolution thresholds under V coincide with $\overline{\theta}_1$ and $\overline{\theta}_2$, and

(ii) the net ex-post surplus from efficiently resolving the partnership dispute is no greater than K,

the partnership dispute (r_1, F_1, F_2, V) cannot be resolved efficiently.

Proof. Using equation (B.14) derived for the net expected utility, the largest ex-ante fee that partner 1 is willing to pay for participation can be bounded from above as follows:

$$\begin{aligned}\min_{\theta_1} U_1(\theta_1; r_1, V) &= \min_{\theta_1} \Big(r_1 \mathbb{E}_{\theta_2} \operatorname{Surp}(\theta_1, \theta_2; V) + r_2 \mathbb{E}_{\theta_2} \operatorname{Impr}_1(\theta_1, \theta_2, V) \Big) \\ &\leqq r_1 \max_{\theta_1, \theta_2} \operatorname{Surp}(\theta_1, \theta_2; V) + r_2 \min_{\theta_1} \mathbb{E}_{\theta_2} \operatorname{Impr}_1(\theta_1, \theta_2, V).\end{aligned}$$

Similarly,

$$\min_{\theta_2} U_2(\theta_2; r_2, V) \leqq r_2 \max_{\theta_1, \theta_2} \operatorname{Surp}(\theta_1, \theta_2; V) + r_1 \min_{\theta_2} \mathbb{E}_{\theta_1} \operatorname{Impr}_2(\theta_1, \theta_2, V).$$

Therefore, the upper bound for the lump-sum fee is

$$\begin{aligned}\mathcal{L}(r_1, r_2, V) &\doteq \min_{\theta_1} U_1(\theta_1; r_1, V) + \min_{\theta_2} U_2(\theta_2; r_2, V) \\ &\leqq \max_{\theta_1, \theta_2} \operatorname{Surp}(\theta_1, \theta_2; V) + r_2 \min_{\theta_1} \mathbb{E}_{\theta_2} \operatorname{Impr}_1(\theta_1, \theta_2, V) + r_1 \min_{\theta_2} \mathbb{E}_{\theta_1} \operatorname{Impr}_2(\theta_1, \theta_2, V).\end{aligned}$$

The expected subsidy can also be rewritten using the above defined Impr_1 and Impr_2 functions:

$$\begin{aligned}\mathcal{S}(r_1, r_2, V) &= r_2 \int_0^1 \int_0^{\underline{\theta}_2(\theta_1)} \theta_1 - \overline{\theta}_1(\theta_2) \, \mathrm{d}F_2(\theta_2) \, \mathrm{d}F_1(\theta_1) \\ &\quad + r_1 \int_0^1 \int_0^{\underline{\theta}_1(\theta_2)} \theta_2 - \overline{\theta}_2(\theta_1) \, \mathrm{d}F_1(\theta_1) \, \mathrm{d}F_2(\theta_2) \\ &= r_2 \mathbb{E}_{\theta_1} \mathbb{E}_{\theta_2} \operatorname{Impr}_1(\theta_1, \theta_2, V) + r_1 \mathbb{E}_{\theta_2} \mathbb{E}_{\theta_1} \operatorname{Impr}_2(\theta_1, \theta_2, V).\end{aligned}$$

Define the values

$$\delta_1 \doteq \mathbb{E}_{\theta_1}\mathbb{E}_{\theta_2}\,\mathrm{Impr}_1(\theta_1,\theta_2,V) - \min_{\theta_1}\mathbb{E}_{\theta_2}\,\mathrm{Impr}_1(\theta_1,\theta_2,V);$$

$$\delta_2 \doteq \mathbb{E}_{\theta_2}\mathbb{E}_{\theta_1}\,\mathrm{Impr}_2(\theta_1,\theta_2,V) - \min_{\theta_2}\mathbb{E}_{\theta_1}\,\mathrm{Impr}_2(\theta_1,\theta_2,V);$$

$$\delta \doteq \delta_1 \wedge \delta_2.$$

Since $f_1, f_2 > 0$, and $\overline{\theta}_1$ and $\overline{\theta}_2$ are not equal to the constant 0 function, the functions $\theta_1 \mapsto \mathbb{E}_{\theta_2}\,\mathrm{Impr}_1(\theta_1,\theta_2)$ and $\theta_2 \mapsto \mathbb{E}_{\theta_1}\,\mathrm{Impr}_2(\theta_1,\theta_2)$ are strictly increasing. Hence, $\delta_1, \delta_2, \delta > 0$, and the budget surplus can be bounded from above as follows:

$$\begin{aligned}\mathcal{L}(r_1,r_2,V) - \mathcal{S}(r_1,r_2,V) &\leqq \max_{\theta_1,\theta_2}\mathrm{Surp}(\theta_1,\theta_2;V)\\ &\qquad + r_2\min_{\theta_1}\mathbb{E}_{\theta_2}\,\mathrm{Impr}_1(\theta_1,\theta_2,V) + r_1\min_{\theta_2}\mathbb{E}_{\theta_1}\,\mathrm{Impr}_2(\theta_1,\theta_2,V)\\ &\qquad - r_2\mathbb{E}_{\theta_1}\mathbb{E}_{\theta_2}\,\mathrm{Impr}_1(\theta_1,\theta_2,V) - r_1\mathbb{E}_{\theta_2}\mathbb{E}_{\theta_1}\,\mathrm{Impr}_2(\theta_1,\theta_2,V)\\ &= \max_{\theta_1,\theta_2}\mathrm{Surp}(\theta_1,\theta_2;V) - r_2\delta_1 - r_1\delta_2 \leqq \max_{\theta_1,\theta_2}\mathrm{Surp}(\theta_1,\theta_2;V) - \delta.\end{aligned}$$

Note that the functions Impr_1 and Impr_2 depend only on the threshold functions $\overline{\theta}_1$ and $\overline{\theta}_2$, but not on other properties of V. Therefore, the same holds for the number $\delta > 0$. Hence by choosing $K \doteq \delta/2$, for every ownership structure (r_1, r_2), and for every partnership value function V such that (i) the dissolution thresholds under V coincide with $\overline{\theta}_1$ and $\overline{\theta}_2$, and (ii) $\max_{\theta_1,\theta_2}\mathrm{Surp}(\theta_1,\theta_2;V) \leqq K$, the partnership dispute (r_1, F_1, F_2, V) cannot be resolved efficiently. ∎

Appendix C

Appendix for Chapter 3

C.1 Proof of Lemma 3.2

Denote the vertices of $F(\mathbf{x})$ by $\mathbf{u^1}, \dots, \mathbf{u^K}$. The result is trivial if $K = 1$ (i.e., $\mathbf{x}$ is a vertex). Assume that $K > 1$. First, we show that for all $h = 1, \dots, K$, there exists a convex combination $\mathbf{x} = \sum_{k=1}^{K} \alpha_k^h \mathbf{u^k}$ where $\alpha_h^h > 0$. First, $\mathbf{x}$ cannot lie on the relative boundary of $F(\mathbf{x})$ (in that case $\mathbf{x}$ would belong to a lower dimensional face). Since $\mathbf{x}$ is in the relative interior, there must be a vector $\mathbf{z} \in F(\mathbf{x})$ such that $\mathbf{x}$ is in the relative interior of the line segment connecting $\mathbf{u^h}$ and $\mathbf{z}$:[1]

$$\mathbf{x} = \beta_1 \mathbf{u^h} + \beta_2 \mathbf{z}, \quad \text{such that } \beta_1, \beta_2 > 0, \text{ and } \beta_1 + \beta_2 = 1.$$

Moreover, since $\mathbf{z} \in F(\mathbf{x})$, the vector $\mathbf{z}$ can be expressed as a convex combination of the vertices of $F(\mathbf{x})$:

$$\mathbf{z} = \sum_{k=1}^{K} \gamma_k \mathbf{u^k}, \quad \text{such that } \gamma_k \geqq 0 \text{ for every } k, \text{ and } \sum_{k=1}^{K} \gamma_k = 1.$$

[1]I.e., there exists $\lambda > 0$ small enough such that $\mathbf{z} \doteq \mathbf{u^h} + (1+\lambda)(\mathbf{x} - \mathbf{u^h}) \in F(\mathbf{x})$.

The last two expressions imply that

$$\mathbf{x} = \underbrace{(\beta_1 + \beta_2\gamma_h)}_{\alpha_h^h > 0}\mathbf{u^h} + \sum_{k \neq h} \underbrace{\beta_2\gamma_k}_{\alpha_k^h}\mathbf{u^k}$$

is a convex combination with the desired property.

Now, take the following convex combination of such convex combinations for every $h = 1, \ldots, K$:

$$\mathbf{x} = \sum_{h=1}^{K} \frac{1}{K} \sum_{k=1}^{K} \alpha_k^h \mathbf{u^h} = \sum_{k=1}^{K} \sum_{h=1}^{K} \frac{1}{K} \alpha_k^h \mathbf{u^h} = \sum_{k=1}^{K} \frac{\sum_{h=1}^{K} \alpha_k^h}{K} \mathbf{u^h}.$$

This is a convex combination of all of the vertices of $F(\mathbf{x})$ such that all of the coefficients are positive since $\alpha_h^h > 0$ for each h. ■