THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

"Mediators and Bilateral Trade"

Ady Pauzner[i], Ran Eilat[ii]

Discussion Paper No. 17-15

October 2015

Thanks to the The Pinhas Sapir Center for Development at Tel Aviv University for their financial support. We thank Andreas Blume, Eddie Dekel, Kfir Eliaz, Jeffrey Ely, Johannes Horner, Nadav Levi, Stephen Morris and Asher Wolinsky for helpful discussions.

[i] Ady Pauzner - The Eitan Berglas School of Economics, Tel Aviv University, Israel. Email:Pauzner@post.tau.ac.il

[ii] Ran Eilat - The Eitan Berglas School of Economics, Tel Aviv University, Israel. Email:eilatran@post.tau.ac.il

Mediators and Bilateral Trade

Ady Pauzner Ran Eilat

Abstract

A mediator communicates with a buyer and seller to decide whether trade occurs and at what price. His objective is to maximize social surplus. Unlike an optimal mechanism that denies trade in some of the realization in which trade is beneficial, a mediator maximizes surplus also at the interim stage: at each point in time he must choose actions that are optimal given the information he has. We study how mediators optimally communicate with the parties, the tradeoffs they face, and the limitations on what they can achieve. In the case that agents' types are distributed uniformly we show that mediators can do no better than the posted-price outcome.

1 Introduction

It is often the case that bilateral negotiations are assisted by mediators whose goal is to lead the two parties to a desired outcome. Peace negotiations, divorce processes, sale of valuable objects, bargaining between creditors to a distressed business – all these are examples in which mediators often play a crucial role. Indeed, mediators may be motivated by selfish considerations such as their share of the surplus, reputation considerations, own preferences over the outcome etc. It is often the case, however, that along with these motivations they are also sincerely interested in the parties' wellbeing. In this paper we abstract from selfish motives and study "benevolent" mediators whose sole goal is to maximize social surplus. How does such a mediator communicate with the

*We thank Andreas Blume, Eddie Dekel, Kfir Eliaz, Jeffrey Ely, Johannes Horner, Nadav Levi, Stephen Morris and Asher Wolinsky for helpful discussions.

parties and make decisions? To what extent can he help the parties to realize the potential social surplus?

Characterizing the optimal mediation procedure differs from the well-studied problem of designing a mechanism that maximizes the social surplus. While a mechanism designer sets a communication and decision policy and then delegates its execution to a hard-wired mechanism, the mediator is also the one in charge of executing it. Thus, unlike a designer of a mechanism, the mediator is bound by his original preference (maximizing surplus) when he eventually communicates with the parties and decides the outcome. In particular, he cannot commit to act "against" the parties, a commitment that is often used by mechanisms to incentivize the parties to reveal private information truthfully and thereby helps to maximize ex-ante social surplus.

For concreteness, consider the canonical buyer-seller problem (Myerson and Satterthwaite 1983) on which we focus in this paper – a simplified environment that captures a major difficulty in reaching an optimal outcome in the above examples. In this private-values setup, a seller owns an object that a buyer potentially wants to buy. The highest social surplus (first-best) is attained if the parties trade whenever the buyer has the higher valuation. But, unless the problem is trivial, this efficient outcome is unattainable: the agents have too strong incentives to misreports their valuations (the seller upwards and the buyer downwards) in order to obtain better trading terms. Myerson and Satterthwaite characterize the optimal (second-best) mechanism – the one that maximizes the gains from trade among the outcomes that are achievable. This mechanism denies trade in realizations in which the buyer's valuation is (slightly) above the seller's. The commitment to sometimes deny trade that is known to be beneficial weakens the agents' incentives to misreport and leads to a higher social surplus on average.

A mediator in the bilateral trade environment also strives to maximize the gains from trade. Unlike a mechanism, however, he cannot commit to deny trade in cases that he knows that such trade is beneficial. Therefore, his only way to credibly deny trade is to know less. He restricts the preciseness of the information he holds by communicating with the agents in a coarse language that pools many types to the same report. By doing so the mediator couples realizations in which trade is beneficial with realizations in which it is not to one information set. This allows him to credibly deny trade. Thus, in contrast to a mechanism for which information is always helpful (as it allows for better allocations), a mediator faces a tradeoff - possessing finer information also reduces the set of outcomes to which he can credibly commit.

The surplus achieved by the optimal mechanism is of course an upper bound on that achievable by a mediator. A closer look shows that it is, in fact, strictly unattainable by a mediator. For example, if the valuations of both agents are uniformly distributed over the interval $[0, 1]$, trade in the optimal mechanism takes place only when the buyer's valuation exceeds the seller's valuation by more than 0.25. To implement such a trading rule, which responds so finely to the agents' types, one must be able to separate all the types of the buyer above 0.25 and all types of the seller below 0.75.[1] Having such fine information, however, the mediator will not be able to credibly deny trade when, for example, the buyer's type is 0.7 and the seller's type is 0.6. Thus, he is not able to implement the optimal trading rule.

A lower bound on the surplus that a mediator can achieve is the posted-price outcome. Here there is a fixed price such that trade occurs whenever the buyer's valuation is above that price and the seller's is below. To implement this "one price fits all" rule the mediator employs a very coarse language, that splits the types of each agent to two – those below and above the posted price. It should be noted, however, that such an outcome can be achieved also without a mediator: a social convention that trade takes place only at that price will bring to the negotiation table only types that will accept it.

Within the range between the two bounds, what is then the social surplus that a mediator can help the parties to realize? To what extent can he adapt the trading decision and price to the specific realization of the parties' valuations? How does the mediator resolve the tension between the need for finer information to implement a better outcome and his inability to commit if his information is too fine?

We begin our analysis by considering a mediator who is restricted to one period of communication with the parties. The mediator chooses the set of messages that is available to each agent and specifies the outcome that will follow each possible profile of reports. This outcome must be credible with respect to the beliefs he has in equilibrium, that is, it must maximize the expected social surplus conditional on the parties' reports. To analyze this 3-player game, we reinterpret the mediator's role in this game as a mechanism with a credibility constraint. Even though the revelation principle does not hold in this context, this allows us to employ powerful tools from the mechanism design theory. We prove that it is without loss to assume away equilibria of mixed strategies. We then show and prove several properties that an optimal mediator must satisfy (e.g.

[1] This statement is true under the assumption that there is only one stage of communication in which both agents report simultaneously, which corresponds to the first part of our analysis. It extends, however, also to the case of more than one stage of communication which we analyze in the second part.

being monotone, deterministic, non-redundant). Given these properties we develop a new and simple technique (which is applicable for monotone mechanisms in general) for determining whether budget balance can hold for a given allocation rule. We then show that, under the assumption of uniform distributions, the optimal mediator implements a simple posted price, i.e., that a mediator cannot improve on the lower bound of what agents can achieve on their own.

In the second part we extend the analysis to the multi-period case. The mediator then chooses a communication protocol and outcomes under the restriction of sequential credibility: at each point in time, given the information at hand, the continuation protocol must be optimal (among all those that satisfy the same property in the sequel). In this part we impose a natural behavioral restriction on the communication vocabulary of the mediator and on the behavior of the agents. This restriction makes the beliefs of the players in the equilibrium of the multi-period game tractable and allows us to analyze properties of the game. We then show that while multi-period communication protocols indeed give the mediator more flexibility in the way he learns information (it allows, for example, for non-monotone trading schemes to arise), under the assumption of uniform distributions the optimal mediator can still do no better than a simple posted price.

1.1 Related Literature

Our work relates to the literature on strategic information transmission as the parties' messages in our mediation model are "cheap-talk": they affect payoffs only indirectly, through the actions of the mediator that depend on the information he learns from these messages. The influential model of cheap-talk games by Crawford and Sobel (1982) shows that information possessed by the sender fails to be fully communicated to the decision maker (receiver) if their interests are not perfectly aligned. (In our model the mediator's interest does not coincide with that of one party because he also cares for the other party.) Equilibrium then involves pooling of senders of close-by types who report the same message. See Sobel (2013) for a recent survey of the huge literature that followed this work.

The paper is also related to the extensive literature on the role of third parties in exchanging information between individuals in cases of conflict. Broadly put, the role of third parties (sometimes referred to as communication devices) is to filter the information transferred between the parties, and thus ease the reluctance of each party to reveal private information. In early contributions, Forges (1986) and Myerson (1986) show that third parties can expand the set of outcomes in these interactions. Gibbons (1988) analyzes the equilibria of two types of arbitration methods when the arbitrator cannot commit to an outcome. In contrast to our work, in the model of Gibbons (1988) the parties have symmetric information and their preferences are independent of the state of the world. For the type of arbitration in which messages are cheap-talk (as in our model) Gibbons (1988) shows the existence of separating equilibria in which the arbitrator extracts all the information from the parties.

In a more recent contribution Goltsman, Horner, Pavlov and Squintani (2009) characterized the optimal structure of dispute resolution institutions within the framework of Crawford and Sobel (1982). They compared the optimal outcomes under arbitration, mediation and negotiation that are designed to maximize the ex-ante welfare of the decision-maker. In contrast to our work in Goltsman et al (2009) there is only one informed party, and the designer can commit to its course of actions. Also, the term mediator is used in their work only for the case of non-binding outcomes whereas in our work the mediator's decision can be either binding or non-binding.[2] Also within the one-sender cheap-talk framework, Blume, Board and Kawamura (2007) study a model in which adding noise to the communication, for example by a mediator, can improve welfare, and Ivanov (2010) studies an environment in which communication is mediated by a strategic third party that can be chosen by the principal.

The role of the designer's commitment power, or the effects of its absence, has also received a considerable amount of attention in the economic literature. In the context of implementation problems with complete information, Baliga, Corchon and Sjostrom (1997) consider a utility maximizing designer that takes an action after all other agents. The designer cannot commit to an outcome function as he might have an incentive to deviate given the agents' reports and the equilibrium that is being played. Their results show that such a designer may achieve sometimes more but sometimes less compared to a standard Nash Implementation. In contrast to our model, in Baliga, Corchon and Sjostrom (1997) the agents have symmetric information, so there always exists an equilibrium in which the agents reveal all the information.[3] Their concern is thus how to knock out the unwanted equilibria. Chakravorty, Corchon and Wilkie (2006) show, in the domain of exchange economies, that if a designer is restricted to off-equilibrium outcomes that must be optimal for at least some preference profiles of agents, then a broad set of social choice functions is

[2] In fact, along most of the analysis we assume binding decisions and discuss the relaxation of this assumption in section 8.2.

[3] Baliga, Corchon and Sjostrom (1997) assume that there are at least 3 agents which are all symmetrically informed, so there always exists a truth-telling equilibrium for the game.

not Nash-implementable.

In the context of contract theory, Bester and Strausz (2001) analyze a principal-agent problem with adverse selection, where the principal has only limited ability to commit to the action he takes following the agent's reports. They show that, when the principal faces one agent whose type space is finite, an extended version of the revelation principal holds (that is, attention can be restricted to contracts in which the message space is the type space). Their results are applicable in several different environments. However, Bester and Strausz (2000) show that the results do not hold for the case of multiple agents.

Skreta (2006) analyzes an environment in which a seller with full bargaining power sequentially offers a trading mechanism to a privately informed buyer at every period along a finite discounted horizon. Skreta (2006) shows that if the seller cannot commit not to offer a new mechanism in case a previous one did not yield trade, then the revenue maximizing strategy is to offer a posted-price mechanism at each period, with prices that vary along time. The analysis in Skreta (2006) assumes very little about the type space of the informed party, and in particular allows for complex (i.e. non-convex) posterior beliefs about its type along the equilibrium path. In another paper, Skreta (2013) analyzes the case of a seller that faces multiple privately informed buyers over a finite discounted horizon and cannot commit not to re-offer the object for sale if it was not sold in previous periods. The rich analysis includes also the possibility for the seller to control the information that the buyers observe over time. In both setups, Skreta (2006) and Skreta (2013), the fact that time is costly and that the seller essentially has full commitment power in the last period plays a crucial role in the solution of the problem.

Vartiainen (2013) analyzes mechanisms that consist of two distinct devices–an information processing device that aggregates reports from the agents and produces a public signal and an implementation device that, given the public signal, determines the outcome of the mechanism. Vartiainen (2013) assumes that the designer can commit to the way the first device operates, but not to the decisions of the second one. Vartiainen (2013) shows that under these assumptions, and in the context of bilateral trade, a careful construction of the mechanism allows implementing the second-best result.

2 The Model

Mediation is a game of three players - a mediator and two agents labeled $i \in \{1,2\}$. Agent 1 (Seller) owns an object that agent 2 (Buyer) potentially wants to buy. Agent i's valuation of the object is denoted v_i and the agents are risk neutral. The mediator's utility is the sum of the agent's utilities, i.e., $v_2 - v_1$ if the object was transferred to the buyer and 0 otherwise. The mediator is risk neutral as well. The valuations of the agents are drawn independently from uniform distributions over $V_i = [\underline{v}_i, \bar{v}_i]$ and are privately known. For non-triviality we assume that the intersection $[\underline{v}_1, \bar{v}_1] \cap [\underline{v}_2, \bar{v}_2]$ is non-empty.

An *environment*, denoted $F = F_1 \times F_2$, is a subset of $V = V_1 \times V_2$ such that $F_1 = \left[\underline{F}_1, \bar{F}_1\right] \subset V_1$ and $F_2 = \left[\underline{F}_2, \bar{F}_2\right] \subset V_2$. A bilateral trade problem is said to be *in* the environment F if it is common knowledge that agent i's type is in F_i for $i \in \{1,2\}$. At the outset the environment of the problem is then $F = V$, but as the game proceeds, it might change.

There are $1 \leq T < \infty$ consecutive periods. T is exogenous to the model and represents the last period at which a decision can be made. At the beginning of period $\tau \in \{1,..,T\}$ the mediator specifies a finite set of possible messages for each agent. Then, each agent may choose to leave the mediation game. If one of the agents leaves, the game terminates without trade or payments. Otherwise the agents simultaneously report messages out of their possible sets. The messages are public and observed by all players. The mediator may then either terminate the game with a decision – whether the object is transferred and at what price, or continue to the next period. At the end of period T the mediator must make a decision.

The solution concept we use for the game is Perfect Bayesian Equilibrium (PBE), with the following refinement: the selected equilibrium induces, at each subgame, the PBE that is preferred by the mediator. This refinement is equivalent to letting the mediator have an additional action at each point he plays, in which he can announce one of the equilibria in the subgame that follows, with the alternative refinement that his recommendation is followed by the agents. For a further discussion of this refinement see section 8.1.

2.1 Mediation Mechanism

As a player in the game the mediator lacks the power to commit to outcomes that follow the agents' actions. This is in contrast to a designer of a mechanism with the same objective (maximizing social surplus). Nevertheless, since both a mediator and a designer of a mechanism face a similar problem – how to effectively elicit private information from the agents – our analysis can benefit from employing tools from the mechanism design literature. We therefore reinterpret the problem as a problem of mechanism design. To that end we replace the mediator in the game (with the refinement above) by a mechanism that is optimal within the set of "credible" mechanisms (to be formally defined later). We now turn to an equivalent description of the mediation process in terms of a mechanism.

A *mediation mechanism* (henceforth MM) is predetermined by the mechanism designer and induces an extensive form game for the buyer and seller. It consists of a game tree of (up to) T periods. In each node of the tree each agent has a set of possible messages. Upon arriving to a node, each agent reports one of his possible messages or chooses the action "leave". The reports are simultaneous and public. If at least one agent chooses to leave, the game is terminated with payoffs of 0 to both. Otherwise they move to the respective node in the next period – unless they are in period T at which, as a function of the reports, the mechanism specifies the final outcome.

Formally, a message of agent i at period τ is denoted $m_{i,\tau}$, and $\mathbf{m}_\tau = (m_{1,\tau}, m_{2,\tau})$ is the profile of messages of both agents at τ. A sequence $(\mathbf{m}_1, ..., \mathbf{m}_\tau)$ starting at period 1 is called a *history* and denoted h.[4] $L(h)$ is the length of history h, that is, the number of elements in h. We denote by H the set of all possible histories. The set H contains the empty history ($\varnothing \in H$) and it has the property that if $(\mathbf{m}_1, ..., \mathbf{m}_K) \in H$ and $L < K$ then $(\mathbf{m}_1, ..., \mathbf{m}_L) \in H$. Denote by $(h, \mathbf{m})$ the sequence that results from the history h followed by $\mathbf{m}$. A history $h \in H$ is *terminal* if there does not exist $\mathbf{m}$ such that $(h, \mathbf{m}) \in H$. Denote by $Z \subseteq H$ the subset of terminal histories in H.

The agents can send messages after every non-terminal history. Denote by $M_i(h)$ the set of messages available for agent i after the non-terminal history $h \in H \backslash Z$ and let $M(h) = M_1(h) \times M_2(h)$. (BY assumption $M_i(h)$ is finite for every $h \in H$ and $i \in \{1, 2\}$.)

A MM is said to be *direct* if the elements in $M_i(\varnothing)$ are a partition of V_i for $i \in \{1, 2\}$, and the elements in $M_i(h)$ are a partition of the last report of agent i in the history h. In a direct MM the messages correspond to subsets of the types' space.

After every terminal history an outcome is determined according to three functions: an allocation function $p : Z \to [0, 1]$ and two payment functions $t_1 : Z \to \mathbb{R}$ and $t_2 : Z \to \mathbb{R}$. The allocation function $p(h)$ is interpreted as the probability that the object is transferred to the buyer after a terminal history $h \in Z$, and $t_i(h)$ is the payment to agent i, which may be negative. Note that

[4] We follow here the formal description of an extensive form game of Osborne and Rubinstein (1994), chapter 11.

there might be non-zero payments even if the object is not transferred. Denote $C = (p, t_1, t_2)$ and let $C(h) = (p(h), t_1(h), t_2(h))$.

A MM must be *ex-post budget balanced*, that is, after every history the net payments to the agents are non-positive:

Property 1 ***(BB)*** *The MM must satisfy ex-post budget balance:*

$$t_1(h) + t_2(h) \leq 0 \text{ for every } h \in Z$$

Finally, a MM with a set of histories H, outcomes C and T periods in the environment F is denoted $G = (H, C, T, F)$. Formally G is a specification of a game form or a mechanism.

Utilities and strategies: The expected utility of agent i of type v_i from the outcome of a terminal history $h \in Z$ is denoted $u_i(v_i, p(h), t_i(h))$ and given by $u_1(v_1, p(h), t_1(h)) = -p(h) \cdot v_1 + t_1(h)$ for the seller and $u_2(v_2, p(h), t_2(h)) = p(h) \cdot v_2 + t_2(h)$ for the buyer.

A behavioral strategy for agent i of type v_i in G is a function $\sigma_i(v_i, h)$ that assigns to every type and every non-terminal history $h \in H \backslash Z$ a distribution of messages in $M_i(h)$. Denote by $\Phi^G_{\sigma_1,\sigma_2}(z \mid h)$ the probability that a terminal history $z \in Z$ will be played in G, given that agents follow σ_1 and σ_2 and given that a non-terminal history h was played so far in G. Denote $\bar{u}^G_i(v_i, \sigma_i, \sigma_{-i} \mid h) = \sum_{z \in Z} \Phi^G_{\sigma_1,\sigma_2}(z \mid h) \cdot u(v_i, p(z), t_i(z))$. This is the interim expected utility of agent i of type v_i from following strategy σ_i in G, after a non-terminal history $h \in H/Z$ and given that agent $-i$ follows the strategy σ_{-i}. When there is no risk of ambiguity we omit the superscript G, and write $\bar{u}_i(v_i, \sigma_i, \sigma_{-i} \mid h)$. Also, if at $h = \varnothing$ (at the beginning of the game) we omit the h and simply write $\Phi^G_{\sigma_1,\sigma_2}(z)$, $\bar{u}^G_i(v_i, \sigma_i, \sigma_{-i})$ etc.

Given a profile of strategies (σ_1, σ_2) that the agents follow in G, we define the *interim social surplus* after a history $h \in H$ in G as follows:

$$W(G, \sigma_1, \sigma_2 \mid h) = \sum_{z \in Z} \Phi^G_{\sigma_1,\sigma_2}(z \mid h) \cdot p(z) \cdot E_{v_1,v_2}[v_2 - v_1 \mid z]$$

W is a weighted sum of the conditional gains from trade in all the possible terminal histories that result from applying the allocation rule p. The weights are the probabilities that each of the terminal histories will be played in G, given that h was played so far and the the agents follow σ_1 and σ_2. The *ex-ante social surplus* is defined as the expected social surplus after the empty history $W_{EA}(G, \sigma_1, \sigma_2) = W(G, \sigma_1, \sigma_2 \mid \varnothing)$.

If G is direct we define a *truthful strategy* for agent i to be a strategy in which after every history $h \in H$, the agent reports the message $m_i \in M_i(h)$ for which $v_i \in m_i$ if such a message exists. We denote the truthful strategy by σ_i^*, and say that agent i *reports truthfully* if he follows σ_i^*. Note that when $T > 1$ a truthful strategy characterizes the behavior of the agent only on the equilibrium path. In these cases there is more than one truthful strategy. Let σ_i^* then denote an arbitrarily chosen one of them. Let $\bar{u}_i^*(v_i \mid h) = \bar{u}_i\left(v_i, \sigma_i^*, \sigma_{-i}^* \mid h\right)$, $W^*(G \mid h) = W(G, \sigma_1^*, \sigma_2^* \mid h)$ and $W_{EA}^*(G) = W_{EA}(G, \sigma_1^*, \sigma_2^*)$.

Agents are assumed to have the outside option not to trade, i.e., a payoff of 0. They stay in the MM as long as their expected continuation payoff is weakly greater than 0.

3 Single Period Mediators

We start by focusing on the case of $T = 1$, i.e. there is only one period of communication before the outcome is determined. We denote a MM with $T = 1$ by $G_1 = (H, C, F)$. The timing of the mediation game is as follows:

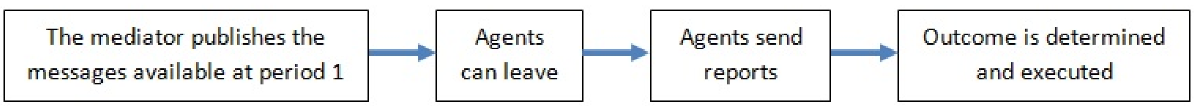


Figure 1: *The timing of a single-period mediation*

Formally, the set of possible histories is given by $H = \{\varnothing\} \cup M(\varnothing)$ where $M(\varnothing) = M_1(\varnothing) \times M_2(\varnothing)$ and $M_i(\varnothing)$ is the set of all the messages available to agent i at the one and only period in the game. For brevity we omit the $\varnothing$ sign and write $M = M_1 \times M_2$. Note that M is the set of terminal histories and is therefore the domain of the outcome functions p, t_1, t_2. We also write $p(m_1, m_2)$ instead of $p((m_1, m_2))$, where $(m_1, m_2) \in M$, and the same for t_i, for $i \in \{1, 2\}$.

In the single-period case, after both agents send their reports, an final outcome is determined. Thus, the two agents play a simultaneous game with incomplete information. A strategy of player i is a function $\sigma_i : V_i \to \Delta(M_1)$ and a Bayesian Nash Equilibrium (BNE) in G is a profile of strategies (σ_1, σ_2) such that each strategy is a best response to the other given the prior beliefs. If such a profile exists we say that G has an equilibrium. Denote by $\mu_i(v_i \mid m_i)$ the posterior distribution of the types of agent i if given that m_i is received in equilibrium, and by $supp(\mu_i(\cdot \mid m_i))$ the support of $\mu_i(\cdot \mid m_i)$.

Consider an equilibrium (σ_1, σ_2) in G_1 and an equilibrium (σ_1', σ_2') in G_1', where both G_1 and G_1' are in the same environment F. We say that the two equilibria are *payoff equivalent* if at the beginning of the game, after each agent has learned his type but before any other information was revealed, each type of each agent expects the same payoff in each of the equilibria. Formally:

Definition 1 *The equilibria* (σ_1, σ_2) *and* (σ_1', σ_2') *in* G_1 *and* G_1'*, respectively, are payoff equivalent if* $\bar{u}_i^G(v_i, \sigma_i, \sigma_{-i}) = \bar{u}_i^{G'}(v_i, \sigma_i', \sigma_{-i}')$ *for every* $v_i \in V_i$ *and* $i \in \{1, 2\}$.

In the single-period case, *credibility* requires that the outcome that follows every profile of reports in equilibrium is the one that maximizes the interim expected social surplus, that is – the expected social surplus conditional on the reports and the information they convey. Formally:

Property 2 (CRED) *The outcome function* $C = (p, t_1, t_2)$ *must be credible with respect to the equilibrium of the MM, i.e.,:*

$$\begin{aligned} & p(m_1, m_2) \cdot \int_{v_2} \int_{v_1} (v_2 - v_1)\, d\mu(v_1 | m_1)\, d\mu(v_2 | m_2) + t_1(m_1, m_2) + t_2(m_1, m_2) \\ \geq \;& p' \cdot \int_{v_2} \int_{v_1} (v_2 - v_1)\, d\mu(v_1 | m_1)\, d\mu(v_2 | m_2) + t_1' + t_2' \end{aligned}$$

for every $p' \in [0, 1]$ *and* $t_1', t_2' \in \mathbb{R}$ *such that* $t_1' + t_2' \leq 0$ *and every* $(m_1, m_2) \in M$*, where* $\mu_i(v_i \mid m_i)$ *are the equilibrium posterior distributions of agents types.*

Two implications follow. First, while ex-post budget balance requires $t_1(m_1, m_2) + t_2(m_1, m_2) \leq 0$ for every profile (m_1, m_2), credibility implies that this condition must holds in equality. We therefore restrict attention to MMs that satisfy ex-post budget balance in strict equality.

Second, credibility implies that without loss of generality we can assume that in the equilibrium of the MM that generates the highest social surplus there are no two types of agent i that are indifferent between two messages available to the agent, for $i \in \{1, 2\}$. This is formally stated in the following lemma:

Lemma 1 *If* $G_1 = (M, C, F)$ *has an equilibrium with respect to which* C *is credible, and in which some* $v_i, v_i' \in V_i$ *are indifferent between reporting some* $m_i, m_i' \in M_i$ *for some* $i \in \{1, 2\}$*, then there exists* $G_1' = (M', C', F)$ *with* $|M_i'| < |M_i|$ *and with an equilibrium which is payoff equivalent and with respect to which* C' *is credible.*

The proof shows that if there are two type of agent i that are indifferent between two messages, it is possible to replace them by one message and define a new MM such that the expected payoff of each type of each agent is unchanged. This is done by exploiting the fact that if two types are indifferent between two messages then the expected probability for trade for each of these two messages must be the same. However it is still necessary to show that the new equilibrium satisfies credibility. It turns out that if the original one satisfied credibility so does the new one. This is since if the two messages, e.g. m_i and m_i', induce the same expected probability for trade for agent i it must be that $p(m_i, m_{-i}) > 0 \Leftrightarrow p(m_i', m_{-i}) > 0$ for each $m_{-i} \in M_{-i}$. Credibility of G implies therefore that the posterior belief regarding whether trade is beneficial or not is the same when m_i or m_i' are reported for every $m_{-i} \in M_{-i}$. This belief, however is preserved for the new message in the new equilibrium and it thus satisfies credibility.

An implication of lemma 1 is that it is possible to restrict attention to MMs with equilibria of pure strategies. When the agents use pure strategies each message can be identified with the set of types that send it in equilibrium (in fact, a message can always be identified with the posterior belief it induces in equilibrium, but when only pure strategies are considered the supports of the beliefs induced by the different messages do not intersect so each message can be identified with the support itself). Thus every $m_i \in M_i$ can be identified with a subset of V_i, such that these subsets are mutually exclusive and their union is V_i. The following corollary is then obtained:

Corollary 1 *Without loss of generality the optimal MM is direct and has an equilibrium in which both agents report truthfully.*

It should be noted that this result is not straightforward in the framework of mechanism design without commitment. In fact, in other environments of mechanism design without commitment it is well known that the opposite is true. For example, Bester and Strauz (2000) show that in a contracting problem when there are multiple agents and the designer cannot fully commit to an allocation function the optimal contract is sometimes achieved when the set of messages is strictly greater than the set of types. In particular this means that some types mix. The structure of our problem, and in particular mediator's objective function rules out the existence of such phenomenon, at least for the case of $T = 1$.

Given M, the truthful strategy is now characterized by $\sigma_i^*(v_i) = \{m_i \in M_i \mid v_i \in m_i\}$, i.e. agent i reports the message $m_i \in M_i$ to which its type v_i "belongs". If the profile (σ_1^*, σ_2^*) constitutes a BNE, we refer to it as a truthful equilibrium. A truthful equilibrium exists if each type of agent i does not expect to gain by deviating from σ_i^*, given that agent $-i$ follows σ_{-i}^*, for $i \in \{1, 2\}$. This is indeed the case if and only if the following *incentive compatibility* condition holds for every $v_i \in V_i$, every possible strategy σ_i and every $i \in \{1, 2\}$:

$$\textbf{(IC)} \quad \bar{u}_i\left(v_i, \sigma_i^*, \sigma_{-i}^*\right) \geq \bar{u}_i\left(v_i, \sigma_i, \sigma_{-i}^*\right)$$

We therefore consider only equilibria that satisfy (IC), those are the truthful BNE.

In a truthful equilibrium of a direct MM the mean expected type of agent i, following the report m_i is $\int_{v_i} v_i d\mu (v_i \mid m_i) = E[v_i \mid v_i \in m_1]$. Credibility then implies (see Property 2) that if $E[v_2 \mid v_2 \in m_2] > E[v_1 \mid v_1 \in m_2]$ then $p(m_1, m_2) = 1$ and if $E[v_2 \mid v_2 \in m_2] < E[v_1 \mid v_1 \in m_2]$ then $p(m_1, m_2) = 0$. The specification of M_1 and M_2 therefore fixes p up to only one degree of freedom (the case that $E[v_2 \mid v_2 \in m_2] = E[v_1 \mid v_1 \in m_2]$). For a truth-telling equilibrium, therefore, a necessary and sufficient condition for credibility to hold is the following condition:

$$\textbf{(CR)} \quad p(m_1, m_2) = \begin{cases} 1 & E[v_2 \mid v_2 \in m_2] > E[v_1 \mid v_1 \in m_1] \\ 0 & E[v_2 \mid v_2 \in m_2] < E[v_1 \mid v_1 \in m_1] \end{cases} \quad \forall m_1 \in M_1, m_2 \in M_2$$

Note that condition (CR) does not fix the value of p when $E[v_2 \mid v_2 \in m_2] = E[v_1 \mid v_1 \in m_1]$.

A *single-period mediator* (henceforth SPM) is defined as follows:

Definition 2 *A single-period mediator (SPM) in the environment F is a direct MM in F for which truth-telling is BNE that satisfies (CR), and (BB) is satisfied with strict equality.*

Let $\mathcal{G}_1(F)$ be the set of all SPMs in the environment F. G_1 is said to be *optimal* in F if it maximizes the ex-ante social surplus within the set $\mathcal{G}_1(F)$:

Definition 3 $G_1 = (M, C, F)$ *is optimal in F if $W^*_{EA}(G_1) \geq W^*_{EA}(G'_1)$ for every $G'_1 \in \mathcal{G}_1(F)$.*

Sections 4 and 6 characterize the optimal Single-Period Mediator.

4 Properties of Optimal Single Period Mediators

Fix a SPM with $G_1 = (M, C, F)$. Let Π_i^F denote the probability distribution of agent i's type in the environment F, with density π_i^F.[5] Then $\bar{p}_i(m_i) = \int_{v_{-i}} p\left(m_i, \sigma^*_{-i}(v_{-i})\right) \pi^F_{-i}(v_{-i})\, dv_{-i}$ is the expected probability that the object is transferred when agent i reports $m_i \in M_i$ and agent $-i$ follows σ^*_{-i}. Similarly, $\bar{t}_i(m_i) = \int_{v_{-i}} t_i\left(m_i, \sigma^*_{-i}(v_{-i})\right) \pi^F_{-i}(v_{-i})\, dv_{-i}$ is the expected payment *to* agent i that reports $m_i \in M_i$, given that agent $-i$ follows σ^*_{-i}. Let $\bar{u}^*_i(v_i) \equiv \bar{u}^*_i(v_i \mid \varnothing)$, this is the payoff of type v_i in the truthful equilibrium, and note that $\bar{u}^*_i(v_1) = -\bar{p}_1(\sigma^*_1(v_1)) \cdot v_1 + \bar{t}_1(\sigma^*_1(v_1))$ and $\bar{u}^*_i(v_2) = \bar{p}_2(\sigma^*_2(v_2)) \cdot v_2 + \bar{t}_2(\sigma^*_2(v_2))$.

[5] In this paper the Π_i^F is always the uniform distribution over F, so Π_i^F and π_i^F are simply $\Pi_i^F(v_i) = \frac{v_i - \underline{F}_i}{\overline{F}_i - \underline{F}_i}$ and $\pi_i^F(v_i) = \frac{1}{\overline{F}_i - \underline{F}_i}$. However, large part of the analysis that follows does not rely on these specific forms.

4.1 Message Sets

The set M is said to be *redundant* if there are two messages that yield the same expected probability of trade for one of the agents in the truthful equilibrium of G_1:

Definition 4 *The set M is redundant in G_1 if there exist $m_i, m_i' \in M_i$ such that $m_i \neq m_i'$ and $\bar{p}_i(m_i) = \bar{p}_i(m_i')$ for some $i \in \{1, 2\}$.*

As an implication of lemma 1 it is without loss of generality to focus on SPMs whose message sets are non-redundant. This is since if M is redundant and $\bar{p}_i(m_i) = \bar{p}_i(m_i')$ for some $m_i, m_i' \in M_i$, condition (IC) implies that $\bar{t}_i(m_i) = \bar{t}_i(m_i')$,[6] so there exist $v_i \in m_i$ and $v_i' \in m_i'$ that are indifferent between reporting m_i and m_i'. We therefore assume that the message sets of the optimal SPM are non-redundant. We proceed with the following lemma that asserts that the set of types that report each message in the truthful equilibrium is convex:

Lemma 2 *If G is an SPM then $\alpha v_i + (1-\alpha) v_i' \in m_i$ for every $m_i \in M_i$, every $v_i, v_i' \in m_i$, every $\alpha \in [0, 1]$ and every $i \in \{1, 2\}$.*

Since in a direct MM the elements $m_i \in M_i$ are a partition of V_i, and since M_i is finite and each m_i is convex, it is possible to enumerate the messages in M_i in ascending order. Let m_i^k denote the k^{th} message in M_i. For a every message $m_i^k \in M_i$, let $\bar{m}_i^k = \sup \left\{v_i \mid v_i \in m_i^k\right\}$ and $\underline{m}_i^k = \inf \left\{v_i \mid v_i \in m_i^k\right\}$, and denote $\left|m_i^k\right| = \bar{m}_i^k - \underline{m}_i^k$ as the length of m_i^k. If $k < l$ we say that m_i^k is *lower than* m_i^l.

4.2 Ex-Ante Budget Balance

Denote by $D(G_1)$ the ex-ante expected net payments to the agents in the truthful equilibrium of G_1:

$$D(G_1) = \int_{v_1} \bar{t}_1(\sigma_1^*(v_1))\, \pi_1^F(v_1)\, dv_1 + \int_{v_2} \bar{t}_2(\sigma_2^*(v_2))\, \pi_2^F(v_2)\, dv_2 \qquad (1)$$

G_1 is said to satisfy *ex-ante budget balance* if $D(G_1) \leq 0$. We refer to $D(G_1)$ as the *deficit* of the SPM G_1.

By definition 2 a SPM satisfies budget balance ex-post (with strict equality). However, it has already been established in the literature (see Borgers and Norman, 2008) that every mechanism with an equilibrium that satisfies budget balance ex-ante has a corresponding mechanism with an

[6] Otherwise, if $\bar{t}_i(m_i') > \bar{t}_i(m_i)$ then $v_i \in m_i$ prefers reporting m_i' rather than m_i.

equilibrium that is payoff equivalent and satisfies budget balance ex-post. This result, which holds for the case of mechanisms continues to hold also in the case of SPMs. To see why note that the translation between a mechanism that is balanced ex-ante and its ex-post equivalent involves only a change in the specification of payments and not in the allocation rule. Since SPMs are mechanisms with a restriction on the allocation rule (the CR condition), but not on payments, the exact argument applies for the case of SPMs. [7] Formally:

Lemma 3 *For every MM with an equilibrium that satisfies budget balance ex-ante there exists a MM which satisfies budget balance ex-post, and the two equilibria are payoff equivalent.*

The proof is standard and omitted. In the analysis that follows we therefore allow a SPM to be balanced only ex-ante, and not ex-post.

4.3 Participation Constraints

Voluntary participation implies that the agents can leave the MM after learning their types and observing the available message sets. When the payoff of the outside option is 0 the agents leave if their type is such that the expected payoff from participation is negative. When analyzing equilibrium one must take into account those types who leave without participating, making the specification of the equilibrium more complex.

In standard mechanism design setups there is an elegant way to avoid this complexity by imposing an *individual rationality* constraint on equilibrium. This constraint requires that the expected payoff of all types of all agents is non-negative. Since incentive compatibility implies that the payoff in equilibrium is increasing in the eagerness for trade of the agent, individual rationality usually reduces to require that one type of each agent has a non-negative payoff. In the bilateral trade environment this is:

$$\textbf{(IR)} \quad u_1^*(\bar{v}_1) \geq 0 \text{ and } u_2^*(\underline{v}_2) \geq 0$$

It is important to notice, however, that the fact that imposing individual rationality is without loss of generality hinges on the fact that the designer has enough degrees of freedom to specify the expected trade probabilities in equilibrium. In particular, the designer can set the expected

[7] In particular, if t_1, t_2 are the payment functions for a SPM with a truthful equlibrium (σ_1^*, σ_2^*) that is balanced ex-ante, the following payment functions make it balanced ex-post: $\hat{t}_i(m_i, m_{-i}) = t_1(m_i, m_{-i}) - \frac{1}{2}\left(d(m_i, m_{-i}) - E_{v_{-i}}[d(m_i, \sigma_{-i}^*(v_{-i}))] + E_{v_i}[d(\sigma_i^*(v_i), m_{-i})]\right)$ where $d(m_1, m_2) = t_1(m_1, m_2) + t_2(m_1, m_2)$.

probabilities of trade to be 0 for the types that are less eager to trade of each agent, i.e. for types $[\hat{v}_1, \bar{v}_1]$ of agent 1 and type $[\underline{v}_2, \hat{v}_2]$ of agent 2, where the designer is *free to choose* $\hat{v}_1 \in V_1$ and $\hat{v}_2 \in V_2$.

Suppose that this was not the case and, for example, the designer was restricted to set $\bar{p}_2(\underline{v}_2) > 0$. This is the case, for example, of any SPM in the environment $[0,1] \times [0.5, 1.5]$ due to the (CR) condition. Imposing $u_2^*(\underline{v}_2) \geq 0$ then implies, by incentive compatibility, that the expected payoff for *all* the types of agent 2 must be *strictly* positive, exhausting the budget balance. On the other hand, if $u_2^*(\underline{v}_2) \geq 0$ is not imposed, then those types who expects negative payoff from participation leave. Their payoff is then 0, rather then positive, making it easier to satisfy budget balance. This can be done, for example, by setting $\bar{t}_2(m_2^1)$ to be extremely negative. Similarly, it is possible to make the highest types of agent 1 leave by setting $\bar{t}_1\left(m_1^{|M_1|}\right)$ to be extremely negative.

It follows that restricting attention to equilibria in which both (IR) and (CR) hold is not without loss of generality. However, as described above, it is always possible to make the lower types of agent 2 and the higher types of agent 1 leave using negative enough payments, regardless of the allocation function. Also, when calculating the generated surplus and budget balance of an SPM, there is no difference between the case that type v_i doesn't trade because $\bar{p}_i(v_i) = 0$ and the case that he quits because the expected payoff is negative.

We therefore continue as follows - we look for the optimal SPM that satisfies (IR), but instead of imposing (CR) we impose a relaxed version of it, which doesn't require it to hold if agent 1 reports $m_1^{|M_1|}$ or agent 2 reports m_2^1. If the "optimal" SPM under the relaxed constraint satisfies (CR) then it is optimal. If not, it means that the allocation rule doesn't satisfy (CR) when agent 1 reports $m_1^{|M_1|}$ or when agent 2 reports m_2^1. (CR) can then be recovered by setting the expected payments to be negative enough when $m_1^{|M_1|}$ or m_2^1 (or both) are reported, such that types $\left[\underline{m}_1^{|M_1|}, \bar{m}_1^{|M_1|}\right]$ of agent 1 and $[\underline{m}_2^1, \bar{m}_2^1]$ of agent 2 would leave. This would not affect the generated social surplus or the budget balance, and the SPM would be the optimal one.

We denote the relaxed condition by (CR-R) and define it formally as follows:

$$\textbf{(CR-R)}\quad p(m_1, m_2) = \begin{cases} 1 & \text{if } E[v_2 \mid v_2 \in m_2] > E[v_1 \mid v_1 \in m_1] \text{ and } m_1 \in M_1 \backslash m_1^{|M_1|} \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{and } m_2 \in M_2 \backslash m_2^1 \\ 0 \text{ or } 1 & \text{if } E[v_2 \mid v_2 \in m_2] > E[v_1 \mid v_1 \in m_1] \text{ and } (m_1 = m_1^{|M_1|} \text{ or } m_2 = m_2^1) \\ 0 & \text{if } E[v_2 \mid v_2 \in m_2] < E[v_1 \mid v_1 \in m_1] \end{cases}$$

(CR-R) is identical to (CR) but doesn't restrict p to be 1 if $E[v_2 \mid v_2 \in m_2] > E[v_1 \mid v_1 \in m_1]$ and $m_1 = m_1^{|M_1|}$ or $m_2 = m_2^1$, i.e. it allows the designer to impose no trade if the buyer (seller) sends his lowest (highest) message, regardless of the posterior distribution in equilibrium. We henceforth focus on equilibria that satisfy (CR-R) and (IR).

4.4 Deterministic Allocation Rule

The condition (CR-R) is weaker than (CR), yet it imposes a strong dependency between the sets M_1 and M_2 and the allocation rule p. In fact M fixes p for all profiles (m_1, m_2) in which $E[v_2 \mid v_2 \in m_2] \neq E[v_1 \mid v_1 \in m_1]$. If the SPM is optimal, however, p is characterized also for the case that $E[v_2 \mid v_2 \in m_2] = E[v_1 \mid v_1 \in m_1]$ and implies that trade doesn't take place. Formally:

Lemma 4 *For an optimal SPM, if $m_1 \in M_1, m_2 \in M_2$ are such that $E[v_2 \mid v_2 \in m_2] = E[v_1 \mid v_1 \in m_1]$, then $p(m_1, m_2) = 0$.*

The intuition is the following - suppose that G satisfies the prefix of the lemma. When the agents report (m_1, m_2) the expected surplus from trade is 0. Allocating the object to the buyer with positive probability, that is $p(m_1, m_2) > 0$, implies that information rents have to be paid to the more eager types of each agent (these are $v_1 \leq \underline{m}_1$ and $v_2 > \bar{m}_2$) in order to support the truthful equilibrium. It is then possible to construct a different SPM, with message sets that are more coarse, that yields a higher ex-ante social surplus and doesn't increase the information rents. This contradicts the optimality of G.

An allocation rule p is said to be *deterministic* if its outcome is either trade or no-trade, but never a stochastic combination between the two:

Definition 5 *An allocation rule p is deterministic if $p(m_1, m_2) \in \{0, 1\}$ for every $m_1 \in M_1, m_2 \in M_2$.*

An SPM is said to be deterministic if its allocation rules is deterministic. An implication of (CR-R) and lemma 4 is that the optimal SPM is deterministic.

In particular, if G is optimal then $p\left(m_1^1, m_2^1\right) \in \{0, 1\}$. Since M is assumed to be non-redundant and due to (CR-R) it follows that $p\left(m_1^k, m_2^k\right) = p\left(m_1^1, m_2^1\right)$ for every k. Furthermore, if $k_1 > k_2$ then $p\left(m_1^{k_1}, m_2^{k_2}\right) = 0$ and if $k_1 < k_2$ then $p\left(m_1^{k_1}, m_2^{k_2}\right) = 1$ for $1 \leq k_1 \leq |M_1|, 1 \leq k_2 \leq |M_2|$. Finally, if $p\left(m_1^1, m_2^1\right) = 0$ then $0 \leq |M_2| - |M_1| \leq 1$, and if $p\left(m_1^1, m_2^1\right) = 1$ then $0 \leq |M_1| - |M_2| \leq 1$.

5 Deterministic Monotone Mechanisms

In this section we temporarily depart from the analysis of SPMs and derive results that apply to a broader set of bilateral trade mechanisms. We keep the notations and definitions as before but allow for a slightly more general setup. We assume that agent i's type is drawn independently from a probability distribution Π_i^F, not necessarily uniform, with support $F_i \subset V_i$ and a density function π_i^F that is bounded and non-negative on F_i.

A mechanism in an environment F, denoted Γ, is a triplet $\Gamma = (M, C, F)$. Without loss of generality we apply the revelation principle and restrict attention to direct revelation mechanisms ($M_i = F_i$) that have a truthful BNE (that is, a BNE that satisfies (IC) as defined above, so that agents report their types truthfully), and individual rationality (IR). For simplicity we focus on mechanisms in which the lowest type of the buyer ($\underline{F}_2$) and highest type of the seller ($\bar{F}_1$) expects to receive the payoff 0, but the results hold (up to straightforward minor changes) if this assumption is relaxed.

In accordance with definition 5, an allocation rule p is said to be *deterministic* if $p(v_1, v_2) \in \{0, 1\}$ for every $v_1 \in F_1, v_2 \in F_2$ and a mechanism is said to be deterministic if its allocation rule is deterministic.

5.1 Deterministic Map of Trade

Fix an arbitrary deterministic allocation rule $p : F \to \{0, 1\}$. Denote $Q = \{(v_1, v_2) \mid p(v_1, v_2) = 1\} \subseteq F$, this is the set of all profiles of types $(v_1, v_2) \in F$ for which trade takes place under p. The set Q is said to be the *deterministic map of trade* (henceforth DMT) induced by p. Figure **??** shows an example of Q projected on the $F_1 \times F_2$ plane when $p(v_1, v_2) = 1$ for $0.2 \leq v_1 \leq v_2 \leq 0.8$ and 0 otherwise. The colored area corresponds to pairs (v_1, v_2) for which $(v_1, v_2) \in Q$.

We say that Q is *monotone* in the environment F if whenever type $v_2 \in F_2$ trade with type $v_1 \in F_1$ under p, that is $p(v_1, v_2) = 1$, then all agent 2's types that are higher than v_2 in F_2 also trade with v_1, and all agent 1's types that are lower than v_1 in F_1 also trade with v_2. We say that Q is *quasi-monotone* in F if there exists $F' \subset F$ in which Q is monotone. Formally:

Definition 6 *Q is monotone in F if $(v_1, v_2) \in Q \implies ((v_1', v_2) \in Q$ and $(v_1, v_2') \in Q)$, for every $v_1, v_1' \in F_1$ such that $v_1' < v_1$ and every $v_2, v_2' \in F_2$ such that $v_2' > v_2$*

Definition 7 *Q is quasi-monotone in F if there exists an environment $F' \subset F$ such that Q is monotone in F'.*

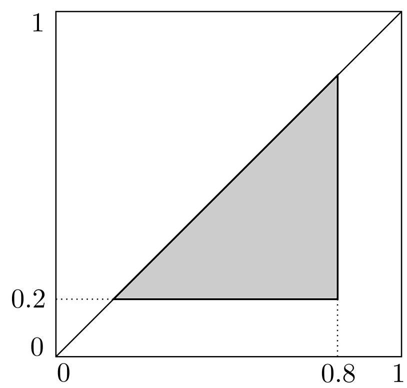


Figure 2: *An example for a deterministic map-of-trade.*

We denote by Q_Γ a DMT induced by the allocation rule of a mechanism Γ. If Q_Γ is monotone in F then Γ is said to be monotone in F.

A DMT Q' is said to be a *sub-map* of a DMT Q if $Q' \subseteq Q$. A set of k sub-maps $\{Q_i\}_{i=1}^k$ is said to *cover* Q if $\bigcup_{i=1}^{k} Q_i = Q$ and if all the sub-maps in $\{Q_i\}_{i=1}^k$ are pair-wise disjoint: $Q_i \cap Q_j = \varnothing$ for every $i \neq j$.

Suppose that Q is quasi-monotone. Let $\eta_H : F \times F_1 \to F_2$ and $\eta_L : F \times F_2 \to F_1$ denote two functions defined as follows:

$$\eta_H(Q, v_2) = \begin{cases} \sup\{v_1 \mid (v_1, v_2) \in Q\} & \text{if } (\underline{F}_1, v_2) \in Q \\ \underline{F}_1 & \text{otherwise} \end{cases}$$

$$\eta_L(Q, v_1) = \begin{cases} \inf\{v_2 \mid (v_1, v_2) \in Q\} & \text{if } \left(v_1, \bar{F}_2\right) \in Q \\ \bar{F}_2 & \text{otherwise} \end{cases}$$

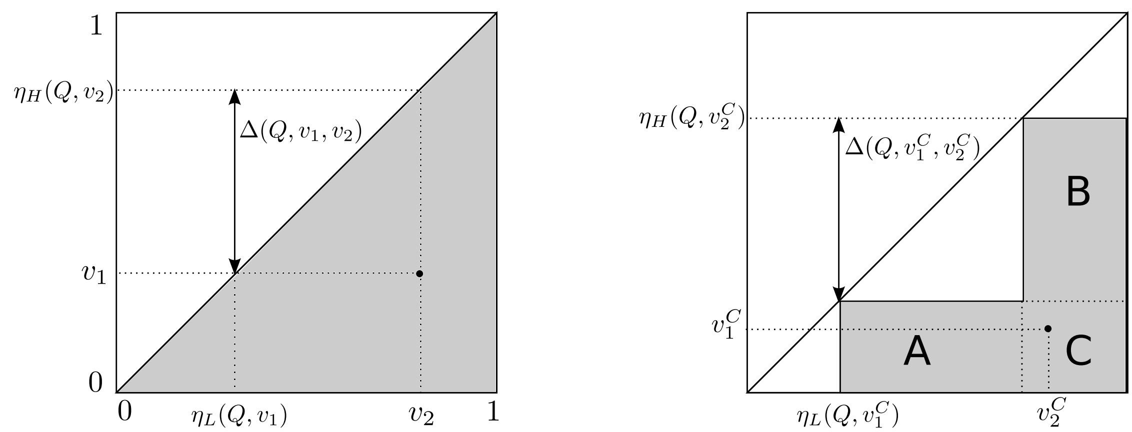
and we also denote:

$$\Delta(Q, v_1, v_2) = \eta_H(Q, v_2) - \eta_L(Q, v_1)$$

The interpretation is as follows: Given Q, $\eta_H(Q, v_2)$ is the highest type of the seller that trades with type $v_2 \in F_2$ of the buyer, or $\underline{F}_1$ if this buyer never trades. Similarly $\eta_L(Q, v_1)$ the lowest type of the buyer that trades with type $v_1 \in F_1$ of the seller, or $\bar{F}_2$ if this seller never trades. $\Delta(Q, v_1, v_2)$ is then the difference between the two, which can be positive or negative. We then define the function $\psi : F \to \mathbb{R}$ as follows:

$$\psi(Q) = \int_{(v_1, v_2) \in Q} \Delta(Q, v_1, v_2) \cdot d\Pi^F(v_1, v_2)$$

The function ψ essentially sums the values of $\Delta(Q, v_1, v_2)$ for all the pairs $(v_1, v_2) \in Q$ weighted by their density.



(a) *Example 1* **(b)** *Example 2*

Figure 3: *Deterministic maps of trade*

5.2 Budget Balance of Deterministic Monotone Mechanisms

Suppose that Γ is deterministic and monotone in some environment F. Assuming that Γ is incentive compatible, its deficit, $D(\Gamma)$, which is the net expected payment to the agents and is given by equation 1, can be rewritten in a simple way with a simple geometric representation:

Proposition 1 *If* $\Gamma = (M, C, F)$ *is a deterministic and monotone mechanism in* F *then* $D(\Gamma) = \psi(Q_\Gamma)$.

Therefore, if Γ satisfies the conditions in the prefix then it is ex-ante budget balanced if and only if $\psi(Q_\Gamma) \leq 0$. Put differently, if $\psi(Q) > 0$ for some monotone Q, then Q could not be induced by *any* ex-ante budget balanced direct revelation mechanism, hence cannot be induced by *any* ex-ante or ex-post budget balanced mechanism.

5.3 Examples

Consider the efficient allocation rule of the bilateral trade problem in the environment $F = [0,1]^2$, that is $p(v_1, v_2) = 1$ if $v_2 \geq v_1$ and 0 otherwise. The DMT that is induced by this allocation rule, denoted Q, is depicted in figure **??**, along with $\eta_L(Q, v_1), \eta_H(Q, v_2)$ and $\Delta(Q, v_1, v_2)$ of a representative pair $(v_1, v_2) \in Q$. Since for every v_1 and v_2 such that $(v_1, v_2) \in Q$, we have $\eta_H(Q, v_2) = v_2$ and $\eta_L(Q, v_1) = v_1$ then $\Delta(Q, v_1, v_2) = (v_2 - v_1) > 0$. It follows that $\psi(Q) = \int_{(v_1,v_2)\in Q} (v_2 - v_1) \cdot d\Pi_i^F(v_1, v_2) > 0$, so Q cannot be induced by any budget balanced mechanism (with voluntary participation). This is a duplication of the famous result of Myerson and Satterthwaite (1983).

For another example consider the DMT Q shown in figure **??**. Q can be divided into three submaps A, B and C. All the pairs (v_1, v_2) in each submap share the same value of $\Delta(Q, v_1, v_2)$. In particular, if $\left(v_1^A, v_2^A\right), \left(v_1^B, v_2^B\right)$ and $\left(v_1^C, v_2^C\right)$ are arbitrarily chosen representative pairs in A, B and C respectively then $\Delta\left(Q, v_1^A, v_2^A\right) = \Delta\left(Q, v_1^B, v_2^B\right) = 0$ and $\Delta\left(Q, v_1^C, v_2^C\right) \equiv \Delta^C > 0$. This quantity appears in the figure and is positive. Calculating $\psi(Q)$ then reduces to calculating Δ^C multiplied by the probability mass of types in C. This is a positive quantity and thus $\psi(Q) > 0$. Q, therefore, cannot be induced by *any* budget balanced mechanism with voluntary participation.

6 The Optimal Single Period Mediator

We turn back now to the analysis of optimal single-period mediator. Suppose that G_1 is a deterministic SPM that satisfies (CR-R). It is therefore also monotone. According to proposition 1, G_1 is ex-ante budget balanced if and only if $\psi\left(Q_{G_1}\right) \leq 0$. Since (CR-R) implies that M fully characterizes p, we can simplify the specification of ψ even further. Denote $\delta\left(m_1^{k_1}, m_2^{k_2}\right) = \bar{m}_1^{k_2-1} - \underline{m}_2^{k_1+1}$ and $\Pr\left(m_1^{k_1}\right) = \Pi_1^F\left(\bar{m}_1^{k_1}\right) - \Pi_1^F\left(\underline{m}_1^{k_1}\right)$, and we then have:

Corollary 2 *if G_1 is deterministic and satisfies (CR-R) then*

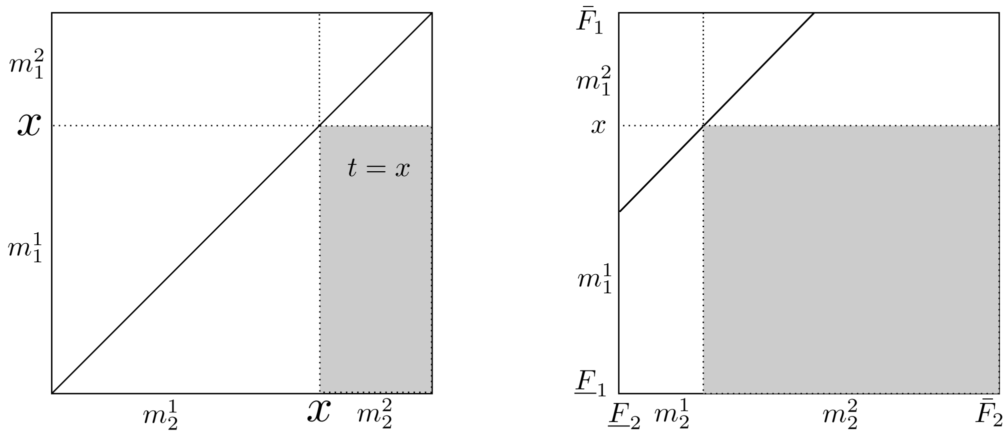
$$\psi\left(Q_G\right) = \begin{cases} \sum_{k_1=1}^{|M_1|} \sum_{k_2=k_1+1}^{|M_2|} \left[\delta\left(m_1^{k_1}, m_2^{k_2}\right) \cdot \Pr\left(m_1^{k_1}\right) \cdot \Pr\left(m_2^{k_2}\right)\right] & \text{if} \quad p\left(m_1^1, m_2^1\right) = 0 \\ \sum_{k_1=1}^{|M_1|} \sum_{k_2=k_1}^{|M_2|} \left[\delta\left(m_1^{k_1}, m_2^{k_2}\right) \cdot \Pr\left(m_1^{k_1}\right) \cdot \Pr\left(m_2^{k_2}\right)\right] & \text{if} \quad p\left(m_1^1, m_2^1\right) = 1 \end{cases}$$

Therefore, the optimal SPM, which is deterministic and satisfies (CR-R), must also satisfy $\psi(Q_G) = 0$ in order to be ex-ante balanced, where $\psi(Q_G)$ is as given by corollary 2.

6.1 Implementing a Posted Price

A very simple type of SPM, though not apriori optimal, is the SPM that *implements some posted price* x. G_1 implements the posted price $x \in F_1 \cap F_2$ in the environment F if it publishes the price x and asks the buyer and the seller to report if their type is below or above x. Formally, $M_i = \left\{m_i^1 = \left[\underline{F}_i, x\right], m_i^2 = \left[x, \bar{F}_i\right]\right\}$ for $i \in \{1, 2\}$. Trade takes place if and only if the seller reports m_1^1 (below x) and the buyer reports m_2^2 (above x). An equivalent definition is:

Definition 8 *G implements a posted price $x \in F_1 \cap F_2$ in the environment F if $Q_G = \{(v_1, v_2) \mid v_1 \leq x, v_2 \geq x\} \subset F$.*



(a) *Example 1* **(b)** *Example 2*

Figure 4: *Implementation of a posted price*

Figure **??** shows a DMT of a SPM that implements some posted price $x \in [0,1]$ in the environment $F = [0,1]^2$. Trade takes place in realizations of types that correspond to the colored area in the graph. The transfers that support this map are $t_1\left(m_1^1, m_2^2\right) = x, t_2\left(m_1^1, m_2^2\right) = -x$ and otherwise 0. It is easy to verify that: (i) the truthful strategies for the agents constitutes an equilibrium, (ii) the outcomes are credible with respect to the truthful equilibrium, (iii) following every profile of reports the net payments to the agents are zero (note also that $\delta\left(m_1^1, m_2^2\right) = 0$ and thus $\psi\left(Q_G\right) = 0$) and (iv) no type of no agent expects a negative payoff. This simple construction is applicable for every environment in which $\underline{F}_2 \leq \underline{F}_1$ and $\bar{F}_2 \leq \bar{F}_1$.

It is only a bit more challenging to implement a posted price $x \in F_1 \cap F_2$ in an environment F for which $\bar{F}_2 > \bar{F}_1$ or $\underline{F}_2 > \underline{F}_1$ (see figure **??** for the case $\bar{F}_2 > \bar{F}_1$ and $\underline{F}_2 > \underline{F}_1$). Suppose that the later holds. In this case, for every $x \in F_1 \cap F_2$ and $m_1^1 = [\underline{F}_1, x]$ and $m_2^1 = [\underline{F}_2, x]$ we have $E\left[v_2 \mid v_2 \in m_2^1\right] > E\left[v_1 \mid v_1 \in m_1^1\right]$. Therefore, whenever the profile $\left(m_1^1, m_2^1\right)$ is reported in a truthful equilibrium trade is beneficial. The only way to implement x as a posted price (in which there is no trade after $\left(m_1^1, m_2^1\right)$) is to make all the types $v_2 \in m_2^1$ of agent 2 leave without participating. This can be done, for example, by setting a negative payment $t_2\left(m_1^1, m_2^1\right) = -\bar{m}_2^1$ so that all types $v_2 \in m_2^1$ expect negative payoff from participating. Thus in equilibrium all types $v_2 \in m_2^1$ of agent 2 leave without participating and do not trade. An analogue argument shows how to implement a posted price if $\bar{F}_2 > \bar{F}_1$. Therefore

Lemma 5 *For every environment F and every $x \in F_1 \cap F_2$ there exists a SPM that implements* *the posted-price x.*

The proof is identical to the construction above and hence omitted. Note that if $\bar{F}_2 > \bar{F}_1$ or $\underline{F}_2 > \underline{F}_1$, or both, the SPM satisfies (CR-R) but not (CR), i.e. it sets $p\left(m_1^1, m_2^1\right) = 0$ even though $E\left[v_2 \mid v_2 \in m_2^1\right] > E\left[v_1 \mid v_1 \in m_1^1\right]$ or sets $p\left(m_1^2, m_2^2\right) = 0$ even though $E\left[v_2 \mid v_2 \in m_2^2\right] > E\left[v_1 \mid v_1 \in m_1^2\right]$, or both.

6.2 Posted-Price and Optimal Single-Period Mediators

The next lemma drives the main result of this section. It states that the optimal SPM always implements a posted price:

Lemma 6 *If G is an optimal SPM in F then it implements some posted price $x \in F_1 \cap F_2$.*

The proof hinges on a tension between budget balance and the (CR-R) condition that is settled only when the SPM implements a posted price. Figure **??** provides a representative example. Suppose that the DMT shown in the example is Q_G induced by a SPM G. Note that for all profiles of reports (m_1, m_2) we have $\delta(m_1, m_2) > 0$, except for the profile $\left(m_1^2, m_2^3\right)$. Note also that $p\left(m_1^2, m_2^2\right) = 0$ so $E\left[v_2 \mid v_2 \in m_2^2\right] \leq E\left[v_1 \mid v_1 \in m_1^2\right]$ must hold due to (CR-R). A necessary condition for budget balance of G is that $-\delta\left(m_1^2, m_2^3\right) \cdot \left|m_1^2\right| \cdot \left|m_2^3\right| > \delta\left(m_1^1, m_2^3\right) \cdot \left|m_1^1\right| \cdot \left|m_2^3\right|$.[8] Using $\delta\left(m_1^1, m_2^3\right) > \left|m_1^2\right|$ and $\left|m_1^1\right| > \underline{m}_1^2 - \underline{m}_2^2$ we get that $-\delta\left(m_1^2, m_2^3\right) > \left(\underline{m}_1^2 - \underline{m}_2^2\right)$ is a weaker necessary condition for budget balance. Writing $\delta\left(m_1^2, m_2^3\right)$ explicitly and rearranging yields $\bar{m}_2^2 + \underline{m}_2^2 > \bar{m}_1^2 + \underline{m}_1^2$ which implies $E\left[v_2 \mid v_2 \in m_2^2\right] > E\left[v_1 \mid v_1 \in m_1^2\right]$, which is a contradiction. The proof shows how this argument generalizes.

Since every posted price in $F_1 \cap F_2$ can be implemented it follows that the optimal SPM implements the posted price that yields the highest expected social surplus. The optimal SPM in the environment $F = \left(\left[\underline{F}_1, \bar{F}_1\right], \left[\underline{F}_2, \bar{F}_2\right]\right)$ is therefore characterized as follows:

Proposition 2 *the optimal SPM in the environment F implements the following posted price:*

$$x = \begin{cases} \underline{F}_2 & \text{if} \quad \frac{\underline{F}_1 + \bar{F}_2}{2} < \underline{F}_2 \\ \frac{\underline{F}_1 + \bar{F}_2}{2} & \text{if} \quad \underline{F}_2 \leq \frac{\underline{F}_1 + \bar{F}_2}{2} \leq \bar{F}_1 \\ \bar{F}_1 & \text{if} \quad \bar{v}_1 < \frac{\underline{F}_1 + \bar{F}_2}{2} \end{cases}$$

[8] This is the place were we use the assumption that the types are uniformly distributed.

The proof is immediate: from lemma (5) and lemma (6) we have that the optimal SPM in an environment F is the optimal posted price in F. Calculating the posted price that maximizes the expected social surplus is straightforward.

7 Multi-Period Mediators

In this section we consider mediation over many periods (any finite $T > 1$). The mediation game then takes the following schematic form:

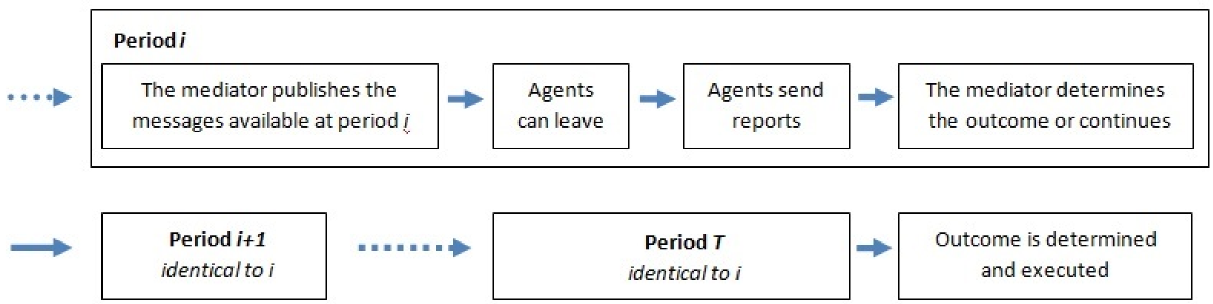


Figure 5: *The timing of mediation games with multiple periods*

Recall that the solution concept we use is Perfect Bayesian Equilibrium with the refinement that at each subgame, the induced PBE is the one preferred by the mediator. This restriction was imposed in order to rule out implausible off-equilibrium "threats" by the agents.

In the multi-period case we replace this refinement with a stronger condition: a restriction on the way the mediator communicates with the agents and on the way the agents behave. We first assume that the language used by the mediator in each node is "simple", in the sense that messages split the type space of each agent into intervals. Second, we assume that the message sets at each node, and the outcome function, are such that truthtelling is an equilibrium in every subgame, and that the agents follow this equilibrium.

More formally, we say that a message set M is a *simple language* in the environment F if M_i constitutes a partition of F_i into intervals. We then assume the following:

Assumption 1: *At each node in the game, the mediator uses a simple language in the environment that is induced by his beliefs regarding the agents' types, truth-telling by the agents is a PBE,*[9] *and the agents report truthfully.*

[9] Such equlibrium always exists as the mediator can offer the agents degenerate message sets (one message for

This restriction is reasonable in our context: the mediator offers a simple set of messages in each stage, and the agents do not have a heavy strategic burden in the analysis the game: they only have to check that reporting their true type is a best response, not to find the equilibrium. More importantly, in cases of multiple equilibria they do not have to coordinate on the chosen one – they have a simple focal point to report truthfully.

Assumption 1 is of course not innocuous. First, it rules out the possibility that the agents randomize between reports that they are indifferent between. Thus, it rules out the possibility of using mixed strategies. Second, it rules out the possibility that the mediator separates types by asking the agents to report messages they are indifferent between (even without randomizing). Indeed, it could (in principle) be the case that by using these techniques the mediator could create information sets that he was not able to create without using them. In particular, by using mixed strategies the mediator could "manipulate" his posterior beliefs to be distributed according to very complex distributions. By imposing this assumption we ensure that the posterior beliefs are tractable. We ensure also that the problem is isomorphic between periods (up to the environment that changes).

7.1 The Mediation Mechanism

We turn again to the interpretation of the game as an interaction between a mechanism and two agents. Note that since there are multiple rounds of interactions between the agents the solution concept is PBE. Since we focus on simple languages we can restrict attention to mechanism which are direct with messages that are always intervals. Also, since by Assumption 1 the truthful equilibrium is always played and messages are intervals, then in equilibrium, after every profile of reports $\mathbf{m} = m_1 \times m_2$, the common belief is that agent i's type is distributed uniformly over the interval m_i.

7.2 Sequential Credibility

In the multi-period case the definition of credibility is more complicated, since the restriction on mechanisms must capture not only the fact that the mediator chooses an optimal outcome given his information, but also that, at each stage, his choice of messages is optimal from that point on. To highlight the difference between the restrictions in the single- and multiple-period cases we refer

each) until the last period, and then implement a posted price.

here to this property as *sequential credibility.*

A formal definition of sequential credibility requires first to define the notion of a *continuation process.* Suppose that $G = (H, C, T, F)$ is a direct MM. For every profile $\mathbf{m} \in M(\varnothing)$, the continuation process of G after $\mathbf{m}$ is defined to be a MM, denoted $G|_{\mathbf{m}} = (H|_{\mathbf{m}}, C|_{\mathbf{m}}, T-1, \mathbf{m})$, where $H|_{\mathbf{m}}$ is the subset of all histories in H that begin with the profile $\mathbf{m}$ as their first element, and $C|_{\mathbf{m}}$ is such that the outcome of every $h \in H|_{\mathbf{m}}$ is the same as its "originating" history in H. Formally: $H|_{\mathbf{m}} = \{h \mid (\mathbf{m}, h) \in H\}$ and $C|_{\mathbf{m}}(h') = C(h)$ for every $h' \in H|_{\mathbf{m}}$, $h \in H$ such that $h' \in h$. The process $G|_{\mathbf{m}}$ is one period shorter than G. Also, given a strategy σ_i in G, let $\sigma_{i|\mathbf{m}}$ denote a strategy that is identical to σ_i after every history in $H|_{\mathbf{m}}$.

In equilibrium, sequential credibility requires that the continuation process after every $\mathbf{m} \in M(\varnothing)$ will maximize the expected social surplus, given that the profile $\mathbf{m}$ was reported at period 1. Recall that, by definition, if the profile (σ_1^*, σ_2^*) constitutes a PBE in G then the profile $\left(\sigma_{1|\mathbf{m}}^*, \sigma_{2|\mathbf{m}}^*\right)$ constitutes a PBE in every continuation process $G|_{\mathbf{m}}$. A mediator is then defined as follows:

Definition 9 *A mediator of $T > 1$ periods in the environment F is a direct MM G in F such that (i) (σ_1^*, σ_2^*) constitutes a PBE in G and (ii) $W_{EA}\left(G|_{\mathbf{m}}, \sigma_{1|\mathbf{m}}^*, \sigma_{2|\mathbf{m}}^*\right) \geq W_{EA}(G')$ for every $G' \in \mathcal{G}_{T-1}(\mathbf{m})$ and every $\mathbf{m} \in M(\varnothing)$, where $\mathcal{G}_{T-1}(\mathbf{m})$ is the set of all mediators of $T-1$ periods in the environment $\mathbf{m}$.*

The definition is recursive in the sense that the definition of a mediator of T periods depends on the definition of a mediator of $T-1$ periods. The definition is complete since the optimal mediator within the set $\mathcal{G}_1(F)$, for every F, was fully characterized in section 6.

A mediator is optimal if it maximizes the ex-ante social surplus from the set of all mediators with T periods in the environment F:

Definition 10 *An mediator G is optimal if $W_{EA}(G) \geq W_{EA}(G')$ for every $G' \in \mathcal{G}_T(F)$.*

Sequential credibility implies that the problem that the mediator faces after every history $h \in H$ is isomorphic to the original problem only in a different environment. The environment of the new problem after history $h = (h', \mathbf{m}) \in H$ is $\mathbf{m} = (m_1, m_2)$. We say that the mediator faces an *easy decision* if, after some history h, his beliefs are such that the buyer's highest type is lower than the seller's lowest type, or that the buyer's lowest type is higher than the seller's highest type, i.e. $m_1 \cap m_2 = \varnothing$. We assume (without loss of generality) that if, after a history h, the mediator faces an easy decision the outcome of the process is realized immediately.

Let H^{T-1} denote the set of histories of length $T-1$, and so $H^{T-1}\backslash Z$ is the set of non-terminal histories of length $T-1$. Suppose that $\mathbf{m}$ is the last element in some history $h \in H^{T-1}\backslash Z$. Since h is not terminal there is no easy decision after h. In period T then the mediator faces the problem of finding the optimal SPM in the environment $\mathbf{m}$. It follows that after every $h = (h', \mathbf{m}) \in H^{T-1}\backslash Z$ the mediator implements the optimal posted price in the environment $\mathbf{m}$.

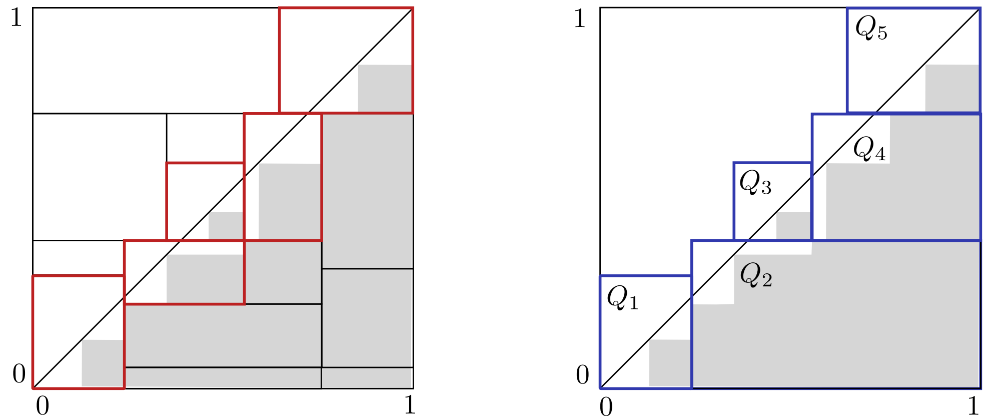
7.3 Deterministic Map of Trade and Budget Balance

If G is an optimal mediator, Q_G is not, apriori, quasi-monotone. We show, however, that it can be covered by a set of quasi-monotone submaps. Formally, we say that a set $\mathcal{Q} \equiv \{Q_k\}_{k=1}^{K}$ is a *quasi-monotone cover of* Q_G of size $K \in \mathbb{N}$ if every $Q \in \mathcal{Q}$ is a quasi-monotone submap of Q_G and $\mathcal{Q}$ covers Q_G. The following lemma asserts that for every mediator G there exists a quasi-monotone cover of Q_G with the property that every submap Q in this cover satisfies $\psi(Q) \geq 0$. Such a cover is denoted $\mathcal{Q}^*(G)$ (and is not necessarily unique). Moreover, if the mediator's non terminal histories are not degenerate, in the sense that at least one of the agents have more than one possible report in periods 1 to $T-1$, then $\psi(Q) > 0$ for at least one of the elements $Q \in \mathcal{Q}^*(G)$.

Lemma 7 *For every mediator $G = (H, C, T, F)$ the DMT Q_G has a quasi-monotone cover $\mathcal{Q}^*(G)$ such that $\psi(Q) \geq 0$ for every $Q \in \mathcal{Q}^*(G)$. If, in addition, there exist $h, h' \in H\backslash Z$ such $L(h) = L(h')$ and $h \neq h'$ then $\psi(Q) > 0$ for at least one element $Q \in \mathcal{Q}^*(G)$.*

The proof shows a construction of $\mathcal{Q}^*(G)$. To see the intuition consider a mediator in the environment $[0,1]^2$ and denote it by G. Since G is deterministic then Q_G is well defined. Figure **??** shows an arbitrary example for such Q_G. note that it is not quasi-monotone. The rectangles in the figure correspond to the environments that the mediator is facing at the beginning of the last period of all the terminal histories $h \in Z$. The union of these rectangles covers $[0,1]^2$. Rectangles that are crossed by the diagonal (the colored ones) correspond to the environments at the beginning of period T for histories $h \in H^{T-1}\backslash Z$. Sequential credibility implies that in all these environments the optimal posted price is implemented at T.

We can then collect the rectangles in the way that is shown in figure **??**, and denote each of these collections in an arbitrary order by Q_k, where $1 \leq k \leq \left|H^{T-1}\backslash Z\right|$. Note that each of these collections contains exactly one rectangle that is crossed by the diagonal, and that a collection can also be a singleton, as in the case of Q_3. Each Q_k is a submap of Q_G, the set $\{Q_k\}_{k=1}^{\left|H^{T-1}\backslash Z\right|}$ covers



(a) *Decomposition to environments*　　**(b)** *A quasi-monotone cover of* Q

Figure 6: *Deterministic map-of-trade of a multi-period mediator*

Q_G and each $Q_{G'}$ is quasi-monotone, thus $\mathcal{Q} = \{Q_k\}_{k=1}^{|H^{T-1}\backslash Z|}$ is a quasi monotone cover of G. It remains to show that $\psi(Q_k) \geq 0$ for each Q_k.

The following lemma establishes the link between the existence of a quasi-monotone cover and the budget balance of G:

Lemma 8 *If G is a mediator and $\mathcal{Q}$ is a quasi-monotone cover of Q_G such that $\sum_{Q\in\mathcal{Q}} \psi(Q) > 0$ then $D(G) > 0$.*

It follows immediately from lemma 7 and lemma 8 that if there exist $h, h' \in H\backslash Z$ such $L(h) = L(h')$ and $h \neq h'$ it means that $D(G) > 0$, an so the MM is not budget balanced. It thus follows that no such h and h' exist. This means that all the reports of both agents along the first $T-1$ periods are degenerate. The mediator elicits no information up to the last period. Sequential credibility then implies that at period T the optimal posted price is implemented. The optimal mediator of T periods is therefore characterized by the following proposition:

Proposition 3 *The optimal mediator in the environment F waits $T-1$ periods and implements the optimal posted price in F at period T.*

8 Discussion

8.1 The Equilibrium Refinement

Along the analysis we have imposed a refinement on the PBE which selects, in every subgame, the continuation equilibrium that yields the highest social surplus (the one preferred by the designer). This refinement is required in order to rule out implausible out-of-equilibrium behavior. Such behaviors can be used as "threats" that support sub-optimal actions on the equilibrium path. Thus, such behavior is a sort of commitment device, the existence of which goes against the spirit of the concept of a mediator.

To see why, consider a mediation game with $T = 2$, and assume that the buyer's and seller's types are independently and uniformly distributed over the interval $[0, 1]$. Assume that at period 1 the mediator offers two possible messages to the buyer ($[0, x]$ and $[x, 1]$) and two possible messages to the seller ($[0, y]$ and $[y, 1]$) such that $x > y$, and that the agents report truthfully.

Let $\mathbf{m} \equiv ([y, 1], [0, x])$. That is, $\mathbf{m}$ is one of the possible profiles of reports at period 1. It is straightforward that if $\mathbf{m}$ is reported at period 1 then a social-surplus maximizing mediator would like to extract more information from the parties at period 2. This is since if he doesn't do so, his optimal decision at the end of period 2 would be "no trade" (mean type of the seller is greater than the mean type of the buyer), and the surplus is then 0. On the other hand, there exists an equilibrium for the subgame that begins after $\mathbf{m}$ in which the mediator extracts more information and achieves a positive social-surplus.

There are, however, other equilibria for that subgame. For example, suppose that regardless of what the mediator does, the buyer and seller babble at the period 2 (that is - randomly choose a message). In that case, at the end of period 2 the mediator does not learn any new information and thus the optimal decision is not to trade, and the social-surplus is 0. Note that this is indeed an equilibrium since no agent can credibly convey information to the mediator and change the decision.

It follows that declaring an outcome and stopping the game after $\mathbf{m}$, without acquiring any further information, can be part of equilibrium. This would be the case if the babbling equilibrium is played in the subgames that follow any of the mediator's action after $\mathbf{m}$ is reported.

The refinement rules out this possibility. It implies that the equilibrium that prevails in every subgame is the one that is preferred by the mediator. In particular it implies that if there exists an equilibrium that yields a strictly positive surplus, the babbling equilibrium will not be played.

8.2 Non-Binding Outcomes

Until now we assumed that the outcomes of the MM are binding, that is, once an outcome is determined the agents are committed to comply. In particular this means that the agents can find ex-post that their payoff from trade was negative.

Suppose, alternatively, that the mediator could decide for each outcome whether it is binding or not. If the outcome is binding the agents must comply, otherwise they do so only if their payoff is non-negative. For non-binding outcomes, therefore, trade takes place only if both agents agree. Sequential credibility implies that the mediator makes an outcome credible if an only if such choice increases the social surplus. We retain the assumption that the agents can leave at any stage of the MM *before* the outcomes are determined.

There are two opposing effects for allowing the mediator to specify outcomes as non-binding. On one hand, non-binding outcomes sometimes yield higher (expected) social surplus than binding ones.[10] On the other hand, allowing for non-binding outcomes decreases the mediator's commitment power – sequential credibility rules out sub-optimal binding outcomes that could possibly be part of an equilibrium.

Note also that a non-binding outcome is, in fact, no more than a recommendation for price for trade. Since the agents cannot communicate after the outcome is determined, all the buyer's type above this price and all the seller's type below it agree to trade. The optimal non-binding outcome is, therefore, to set a price that is equal to the optimal posted price that correspond to the mediator's beliefs.

It follows, therefore, that a MM that allows for binding and non-binding outcomes is equivalent to a MM that allows for only binding outcomes, but allows also that some of the terminal histories would be one period longer ($T+1$ periods), under the constraint that in the last period the optimal posted price is implemented. This implies, however, that the map of trade of such MM has the same

[10] To see why consider an example in which after the agents finished reporting (that is - after a terminal history), when time comes to decide the outcome, the common belief if that the seller's type is distributed uniformly on $[0,1]$ and the buyer's type is distributed uniformly on $[0.2,1.2]$. The best binding outcome is to transfer the object, and price doesn't matter. The best non-binding outcome is to propose the agents to trade at price 0.6. Only sellers below 0.6 and buyers above 0.6 indeed trade. The expected social surplus of the binding outcome is 0.2 and of the non-binding one 0.216. Note, however, that if the buyer was distributed uniformly on $[0.5,1.5]$ the best binding outcome – trade, price doesn't matter – yields an expected surplus of 0.5, while the best non-binding outcome – setting the price 0.75 – yields the expected surplus of 0.42.

properties of the map of trade of a MM with binding outcomes. In particular all the arguments and techniques that were used for the proofs of lemma 7 and lemma 8 still apply. Therefore, allowing the mediator to chose whether an outcome is binding or not does improve the social surplus generated by the optimal mediator.

9 Appendix - Proofs

9.1 Proof of Lemma 1

Suppose that $G_1 = (M, C, F)$ has an equilibrium in which the agents report according to some profile (σ_1, σ_2). Let $q_i(m_i)$ denote the probability that agent i reports m_i in equilibrium. We show the proof for the case that the seller has two types that are indifferent between two messages. A similar proof applies for the buyer.

Suppose also that types $v_1, w_1 \in V_1$ of the seller, with $v_1 \neq w_1$, are indifferent between reporting $m_1, m_1' \in M_1$, with $m_1 \neq m_1'$ and $q(m_1) > 0$ and $q(m_1') > 0$. Therefore:

$$\sum_{m_2 \in M_2} q_2(m_2) \cdot (-p(m_1, m_2) \cdot v_1 + t_1(m_1, m_2)) = \sum_{m_2 \in M_2} q_2(m_2) \cdot \left(-p\left(m_1', m_2\right) \cdot v_1 + t_1\left(m_1', m_2\right)\right)$$

$$\sum_{m_2 \in M_2} q_2(m_2) \cdot (-p(m_1, m_2) \cdot w_1 + t_1(m_1, m_2)) = \sum_{m_2 \in M_2} q_2(m_2) \cdot \left(-p\left(m_1', m_2\right) \cdot w_1 + t_1\left(m_1', m_2\right)\right)$$

It follows that $\sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1, m_2) = \sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1', m_2)$, so the expected probability for trade of the seller is identical when reporting m_1 or m_1'. Denote this probability by λ_p. Thus also $\sum_{m_2 \in M_2} q_2(m_2) \cdot t_1(m_1, m_2) = \sum_{m_2 \in M_2} q_2(m_2) \cdot t(m_1', m_2)$, so the expected payment to the seller is also identical when reporting m_1 and m_1'. Denote it by λ_t.

Now consider an alternative MM, denoted $G' = (M', C', F)$, which is identical to G, only that m_1 and m_1' are replaced by one message denoted α. The allocation rule of G' is $p'(\alpha, m_2) = \frac{q_1(m_1) \cdot p(m_1, m_2) + q_1\left(m_1'\right) \cdot p\left(m_1', m_2\right)}{q_1(m_1) + q_1\left(m_1'\right)}$, and otherwise identical to p. The payments are $t_i'(\alpha, m_2) = \frac{q_1(m_1) \cdot t(m_1, m_2) + q_1\left(m_1'\right) \cdot t\left(m_1', m_2\right)}{q_1(m_1) + q_1\left(m_1'\right)}$ and are otherwise identical to t_i.

The pair of strategies (σ_1', σ_2) in which the buyer follows the same strategy as in G, and the seller follows a strategy which is identical to σ_1, but reports α instead of m_1 and m_1', is an equilibrium in G'. To see why note first that the buyer's incentives haven't changed. For every $m_2 \in M_2$ the expected probability for trade is the same in both cases, and so does the expected payment. Thus, if σ_2 is a best reply to σ_1 in G then it is also a best reply to σ_1' in G'.

For the seller, the expected probability for trade is also the same in both equilibria, and so does the expected payment. This is since:

$$\begin{aligned}\sum_{m_2 \in M_2} q_2(m_2) \cdot p'(\alpha, m_2) &= \sum_{m_2 \in M_2} q_2(m_2) \cdot \frac{q_1(m_1) \cdot p(m_1, m_2) + q_1(m_1') \cdot p(m_1', m_2)}{q_1(m_1) + q_1(m_1')} = \\ &= \frac{q_1(m_1) \cdot \left(\sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1, m_2)\right)}{q_1(m_1) + q_1(m_1')} \\ &\quad + \frac{q_1(m_1') \cdot \left(\sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1', m_2)\right)}{q_1(m_1) + q_1(m_1')} \\ &= \lambda_p\end{aligned}$$

and the same calculation shows that $\sum_{m_2 \in M_2} q_2(m_2) \cdot t(\alpha, m_2) = \lambda_t$.

Finally, we have to show that C' is credible with respect to the equilibrium (σ_1, σ_2). Let μ_i denote the posterior distribution of agent i's type in the equilibrium (σ_1, σ_2) of G, and μ_i' denote the posterior distribution according to the equilibrium (σ_1', σ_2) of G'. If $\int_{v_1} v_1 d\mu(v_1|m_1) = \int_{v_1} v_1 d\mu(v_1|m_1') \equiv \lambda_E$ then $\int_{v_1} v_1 d\mu'(v_1|\alpha) = \lambda_E$. Thus, if C is credible w.r.t (σ_1, σ_2) in G, then C' is credible w.r.t (σ_1', σ_2) in G'.

Otherwise, assume that $\int_{v_1} v_1 d\mu(v_1|m_1) > \int_{v_1} v_1 d\mu(v_1|m_1')$ (the proof for the opposite case is identical). Credibility of C then implies $p(m_1, m_2) \leq p(m_1', m_2)$ for every $m_2 \in M_2$. However, since $\sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1, m_2) = \sum_{m_2 \in M_2} q_2(m_2) \cdot p(m_1', m_2)$ it must be that $p(m_1, m_2) = p(m_1', m_2) = p'(\alpha, m_2)$ for every $m_2 \in M_2$. Credibility of C with respect to the equilibrium (σ_1, σ_2), and the fact that $p(m_1, m_2) = p(m_1', m_2)$, imply that the signs of $\int_{v_2} \int_{v_1} (v_2 - v_1) d\mu(v_1|m_1) d\mu(v_2|m_2)$ and $\int_{v_2} \int_{v_1} (v_2 - v_1) d\mu(v_1|m_1') d\mu(v_2|m_2)$ are identical for every $m_2 \in M_2$. Since $\int_{v_1} v_1 d\mu(v_1|m_1) \geq \int_{v_1} v_1 d\mu'(v_1|\alpha) \geq \int_{v_1} v_1 d\mu(v_1|m_1')$ then it is identical also to the sign of $\int_{v_2} \int_{v_1} (v_2 - v_1) d\mu'(v_1|m_1) d\mu(v_2|m_2)$. Tis sign is therefore consistent with the value of $p(\alpha, m_2)$, and C' is credible with respect to the equilibrium (σ_1', σ_2) in G'.

9.2 Proof of Lemma 2

Let G be a SPM. We show the proof for the case of $i = 2$. The case of $i = 1$ is similar and omitted. For every $m_2 \in M_2$ and $v_2, v_2' \in m_i$ the IC condition implies:

$$\begin{aligned} v_2 \cdot \bar{p}_2(m_2) + \bar{t}_2(m_2) &\geq v_2 \cdot \bar{p}_2(\hat{m}_2) + \bar{t}_2(\hat{m}_2) \\ v_2' \cdot \bar{p}_2(m_2) + \bar{t}_2(m_2) &\geq v_2' \cdot \bar{p}_2(\hat{m}_2) + \bar{t}_2(\hat{m}_2) \end{aligned}$$

for every $\hat{m}_2 \in M_2$ such that $\hat{m}_2 \neq m_2$. Since v_2 and v_2' are not both indifferent between reporting m_2 and $\hat{m}_2$ (see lemma 1) then at least one of the above inequalities is strict. Thus, for every $\alpha \in (0,1)$ we have:

$$\left(\alpha v_2 + (1-\alpha) v_2'\right) \cdot \bar{p}_2(m_2) + \bar{t}_2(m_2) > \left(\alpha v_2 + (1-\alpha) v_2'\right) \cdot \bar{p}_2(\hat{m}_2) + \bar{t}_2(\hat{m}_2)$$

so type $(\alpha v_2 + (1-\alpha) v_2') \in V_2$ finds it optimal to report m_2, or equivalently $(\alpha v_2 + (1-\alpha) v_2') \in m_2$.

9.3 Proof of Lemma 4

Suppose $G = (M, C, F)$ is a SPM and there exist $m_1^l \in M_1, m_2^k \in M_2$ such that $E\left[v_2 \mid v_2 \in m_2^k\right] = E\left[v_1 \mid v_1 \in m_1^l\right]$ and $p\left(m_1^l, m_2^k\right) = \lambda$ for some $\lambda > 0$. The proof shows how a change in the specification of G increases the ex-ante expected social surplus from trade, without increasing the deficit $D(G)$ (see equation 1), and without violating the incentive compatibility or credibility of G. For brevity we denote $a_1 = \underline{m}_1^l$, $a_2 = \underline{m}_2^k$, $b_1 = \bar{m}_1^l$, $b_2 = \bar{m}_2^k$ and $\gamma_2 = \lambda a_2 + (1-\lambda) b_2$. Without loss of generality we assume that $a_2 \geq a_1$ (otherwise the proof is similar but with the roles of the buyer and seller reversed).

Consider the following change in G. We replace the message $m_2^k \in M_2$ with the two new messages $m_2^{kA} = [a_2, \gamma_2]$ and $m_2^{kB} = [\gamma, b_2]$ and define the outcomes for the cases that the buyer reports m_2^{kA} and m_2^{kB} as follows:

- $p\left(m_1^l, m_2^{kA}\right) = 0$ and $p\left(m_1^{l'}, m_2^{kA}\right) = p\left(m_1^{l'}, m_2^{kB}\right) = 0$ for every $l' > l$.
- $p\left(m_1^l, m_2^{kB}\right) = 1$ and $p\left(m_1^{l''}, m_2^{kA}\right) = p\left(m_1^{l''}, m_2^{kB}\right) = 1$ for every $l'' < l$.
- $t_2\left(m_1^l, m_2^{kA}\right) = t_2\left(m_1^l, m_2^k\right) + \lambda a_2$
- $t_2\left(m_1^l, m_2^{kB}\right) = t_2\left(m_1^l, m_2^k\right) - (1-\lambda) \cdot b_2$
- $t_1(m_1, m_2) = t_1(m_1, m_2) + \lambda(1-\lambda)(b_1 - a_1)(b_2 - a_2)^2$ for every $m_1 \in M_1$ and $m_2 \in M_2$.

For all other cases the outcome functions are not changed (with the interpretation that if the buyer reports m_2^{kA} or m_2^{kB}, the outcome is determines as if he sent m_2^k before the change.

We now show that if the original SPM satisfied (IC), (CR) and (BB), so does the altered one:

Credibility (CR): The allocation $p\left(m_1^l, m_2^{kA}\right) = 0$ is consistent with $E\left[v_2 \mid v_2 \in m_2^{kA}\right] < E\left[v_1 \mid v_1 \in m_1^l\right]$ and $p\left(m_1^l, m_2^{kB}\right) = 1$ is consistent with $E\left[v_2 \mid v_2 \in m_2^{kB}\right] > E\left[v_1 \mid v_1 \in m_1^l\right]$. Since $\bar{m}_2^k \leq E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$[11] then

[11] This follows from the assumption of uniform distributions, since if $\bar{m}_2^k > E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$ then it couldn't

$E\left[v_2 \mid v_2 \in m_2^{kA}\right] \leq E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$ and $E\left[v_2 \mid v_2 \in m_2^{kB}\right] \leq E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$, so $p\left(m_1^{l'}, m_2^{kA}\right) = p\left(m_1^{l'}, m_2^{kB}\right) = 0$ is consistent with (CR) for every $l' > l$. Also, since $\underline{m}_2^k \geq \bar{m}_1^l$ then $E\left[v_2 \mid v_2 \in m_2^{kA}\right] > E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$ and $E\left[v_2 \mid v_2 \in m_2^{kB}\right] > E\left[v_1 \mid v_1 \in m_1^{l}\right]$, and thus $p\left(m_1^{l''}, m_2^{kA}\right) = p\left(m_1^{l''}, m_2^{kB}\right) = 1$ is consistent with (CR) for every $l'' < l$. Therefore (CR) is satisfied.

Incentive Compatibility (IC): The incentives of agent 1 to report truthfully have not changed since. This is since, for every $m_1 \in M_1$, the seller's expected probability of trade has not changed and the expected payment has shifted by a constant.

Regarding agent 2, note that for every $v_2 \in V_2$ the expected payoff from reporting m_2^{kB} is increased (relative to reporting m_2^k) by $(b_1 - a_1)((1-\lambda)v_2 - (1-\lambda)b_2)$ and the expected payoff from reporting m_2^{kA} is increased (relative to reporting m_2^k) by $(b_1 - a_1)(-\lambda v_2 + \lambda a_2)$. It follows that the payoffs of types $b_2 \in V_2$ and $a_2 \in V_2$ from reporting m_2^{kB} and m_2^{kA}, respectively, are unchanged relative to their payoff from reporting m_2^k in the original SPM. Thus type b_2 is indifferent between reporting m_2^{kB} and m_2^{k+1}, and type a_2 is indifferent between reporting m_2^{kA} and m_2^{k-1}. Also, type $\gamma_2 \in V_2$ is indifferent between reporting m_2^{kA} and m_2^{kB}.

If an arbitrary type $v_2 \in V_2$ is indifferent between reporting two messages $m_2^{k'}$ and $m_2^{k''}$, with $k'' > k'$ and $\bar{p}(k'') \geq \bar{p}(k')$, then every type lower than v_2 strictly prefers reporting $m_2^{k'}$ to $m_2^{k''}$, and every type greater than v_2 strictly prefers reporting $m_2^{k''}$ to $m_2^{k'}$. It thus follows that if reporting truthfully was an equilibrium in the original SPM, it is also an equilibrium after the change.

Budget Balance (BB): The expected payments to the buyer are increased by $(b_1 - a_1)(\gamma_2 - a_2)(\lambda \cdot a_2) + (b_1 - a_1)(b_2 - \gamma_2)(-(1-\lambda) \cdot b_2)$, which is, after a simple an manipulation, equal to the (negative) quantity $-\lambda(1-\lambda)(b_1 - a_1)(b_2 - a_2)^2$. The expected payments to the seller are increased by $\lambda(1-\lambda)(b_1 - a_1)(b_2 - a_2)^2$ so the deficit of the SPM is not affected - if the original SPM was budget balanced, it is so also after the change.

It therefore remains to show that the surplus generated by trade is increased. The change in the expected surplus is given by:

$$\begin{aligned} & \int_{a_2}^{\gamma_2} \int_{a_1}^{b_1} (0 - \lambda) \cdot (v_2 - v_1)\, dv_1 dv_2 + \int_{\gamma_2}^{b_2} \int_{a_1}^{b_1} (1 - \lambda) \cdot (v_2 - v_1)\, dv_1 dv_2 \\ = \; & \frac{1}{2}(1-\lambda)\lambda(b_1 - a_1)(a_2 - b_2)^2 \end{aligned}$$

which is a positive quantity.

be that $E\left[v_2 \mid v_2 \in m_2^k\right] = E\left[v_1 \mid v_1 \in m_1^l\right]$. While it is outside the scope of this work, note that for case of other distributions, an extension to the proof could be easily added. If $\bar{m}_2^k > E\left[v_1 \mid v_1 \in m_1^{l+1}\right]$ then setting $p\left(m_1^l, m_2^k\right) = 0$ and slightly decreasing $\bar{m}_2^k$ would yield a new SPM with lower deficit and higher expected social surplus.

9.4 Proof of Proposition 1

Suppose that Γ is an arbitrary mechanism that satisfies (IR) and (IC) in the environment $F = \left[\underline{v}_1^F, \bar{v}_1^F\right] \times \left[\underline{v}_2^F, \bar{v}_2^F\right]$, and $u^*\left(\bar{v}_1^F\right) = u\left(\underline{v}_2^F\right) = 0$ (i.e. the highest type of the seller and the lower type of the buyer expects payoff 0 in the truthful equilibrium). Let p denote the allocation function Γ and let $\bar{p}_i(v_i) = \int_{\underline{v}_{-i}^F}^{\bar{v}_{-i}^F} p(v_1, v_2)\,\pi_{-i}^F(v_{-i})\,dv_{-i}$, this is the expected probability that type v_i trade in the truthful equilibrium. The DMT induced by p is denoted Q_Γ.

Since Γ is incentive compatible, the payoff of type v_1 is given by $\int_{v_1}^{\bar{v}_1^F} \bar{p}_1(v_1)\,dx$ and the payoff of type v_2 is given by $\int_{\underline{v}_2^F}^{v_2} \bar{p}_2(v_2)\,dx$. The expected payment to type v_1 of agent i is therefore given by $v_1 \cdot \bar{p}_1(v_1) + \int_{v_1}^{\bar{v}_1^F} \bar{p}_1(v_1)\,dx$, and the expected payment to type v_2 of agent 2 is given by $-v_2 \cdot \bar{p}_2(v_2) + \int_{\underline{v}_2^F}^{v_2} \bar{p}_2(v_2)\,dx$.

Consider the following two payment functions (regardless of those actually specified in Γ):

$$t_1(v_1, v_2) = \begin{cases} \eta_H(Q_\Gamma, v_2) & \text{if } v_2 > \eta_L(Q_\Gamma, v_1) \\ 0 & \text{otherwise} \end{cases}$$

$$t_2(v_1, v_2) = \begin{cases} -\eta_L(Q_\Gamma, v_1) & \text{if } v_1 < \eta_H(Q_\Gamma, v_2) \\ 0 & \text{otherwise} \end{cases}$$

and define $\bar{t}_i(v_i) = \int_{\underline{v}_{-i}^F}^{\bar{v}_{-i}^F} t_i(v_1, v_2)\,\pi_{-i}^F(v_{-i})\,dv_{-i}$. We first show that the expected payment to type v_i of agent i in Γ is $\bar{t}_i(v_i)$, i.e. we show that $\bar{t}_1(v_1) = v_1 \cdot \bar{p}_1(v_1) + \int_{v_1}^{\bar{v}_1^F} \bar{p}_1(v_1)\,dx$ and $\bar{t}_2(v_2) = -v_2 \cdot \bar{p}_2(v_2) + \int_{\underline{v}_2^F}^{v_2} \bar{p}_2(v_2)\,dx$. For the case of $\bar{t}_2$ note that, by definition, $\bar{t}_2(v_2) = \int_{\underline{v}_1^F}^{\eta_H(v_2)} \left[-\eta_L(Q_\Gamma, v_1)\right] \pi_1^F(v_1)\,dv_1$, and thus:

$$\begin{aligned}
\bar{t}_2(v_2) &= -\Pi_1(\eta_H(Q_\Gamma, v_2)) \cdot v_2 + \int_{\underline{v}_1^F}^{\eta_H(v_2)} \left[v_2 - \eta_L(Q_\Gamma, v_1)\right] \pi_1^F(v_1)\,dv_1 \\
&= -\Pi_1(\eta_H(Q_\Gamma, v_2)) \cdot v_2 + \int_{\underline{v}_1^F}^{\eta_H(v_2)} \int_{\eta_L(Q_\Gamma, v_1)}^{v_2} \pi_1^F(v_1)\,dx dv_1 \\
&= -\Pi_1(\eta_H(Q_\Gamma, v_2)) \cdot v_2 + \int_{\bar{v}_2^F}^{v_2} \int_{\underline{v}_1^F}^{\eta_H(x)} \pi_1^F(v_1)\,dv_1 dx \\
&= -\Pi_1(\eta_H(Q_\Gamma, v_2)) \cdot v_2 + \int_{\underline{v}_2^F}^{v_2} \Pi_1(\eta_H(Q_\Gamma, x))\,dx
\end{aligned}$$

where the third equation follows from the fact that[12]:

$$\begin{aligned}
& \left\{(v_1, v_2) : \underline{v}_1^F \leq v_1 \leq \bar{v}_1^F, \eta_L(Q_\Gamma, v_1) \leq v_2 \leq \bar{v}_2^F\right\} \\
= & \left\{(v_1, v_2) : \underline{v}_2^F \leq v_2 \leq \bar{v}_2^F, \underline{v}_1^F \leq v_1 \leq \eta_H(Q_\Gamma, v_2)\right\}
\end{aligned}$$

[12] See Protter and Morrey (1985)

Since $\Pi_1(\eta_H(Q_\Gamma, v_2)) = \bar{p}_2(v_2)$ then $\bar{t}_2(v_2) = -v_2 \cdot \bar{p}_2(v_2) + \int_{\underline{v}_2^F}^{v_2} \bar{p}_2(x)\,dx$. A similar argument applies for $\bar{t}_1(v_1)$ and is omitted.

In order to show that the deficit of Γ is equal to $\psi(Q_\Gamma)$ we write $D(\Gamma)$ explicitly as the sum of the expected payments to the agents:

$$\begin{aligned}
D(\Gamma) &= \int_{\underline{v}_1^F}^{\bar{v}_1^F} \bar{t}_1(v_1)\,\pi_1^F(v_1)\,dv_1 + \int_{\underline{v}_2^F}^{\bar{v}_2^F} \bar{t}_2(v_2)\,\pi_2^F(v_2)\,dv_2 \\
&= \int_{\underline{v}_1^F}^{\bar{v}_1^F} \left[\int_{\eta_L(v_1)}^{\bar{v}_2^F} \eta_H(Q_\Gamma, v_2)\,\pi_2^F(v_2)\,dv_2\right] \pi_1^F(v_1)\,dv_1 \\
&\quad + \int_{\underline{v}_2^F}^{\bar{v}_2^F} \left[\int_{\underline{v}_1^F}^{\eta_H(v_2)} -\eta_L(Q_\Gamma, v_1)\,\pi_1^F(v_1)\,dv_1\right] \pi_2^F(v_2)\,dv_2 \\
&= \int_{\underline{v}_2^F}^{\bar{v}_2^F} \int_{\underline{v}_1^F}^{\bar{v}_1^F} 1_{(v_1,v_2)\in Q} \cdot [\eta_H(Q_\Gamma, v_2) - \eta_L(Q_\Gamma, v_1)] \cdot \pi_1^F(v_1)\,\pi_1^F(v_2)\,dv_1 dv_2 \\
&= \int_{(v_1,v_2)\in Q_\Gamma,} \Delta(Q_\Gamma, v_1, v_2)\,d\Pi_i^F(v_1, v_2) \\
&= \psi(Q_\Gamma)
\end{aligned}$$

9.5 Proof of Lemma 6

Let G be an optimal SPM, and assume that it *does not* implement a posted price. Along the proof we assume that $p(m_1^1, m_2^1) = 0$, so $p\left(m_1^l, m_2^{l+j}\right) = 1$ for $j \geq 1$ and $p\left(m_1^l, m_2^{l+j}\right) = 0$ for $j \leq 0$. The proof for the case of $p(m_1^1, m_2^1) = 1$ is similar and involves only an offset of indices.

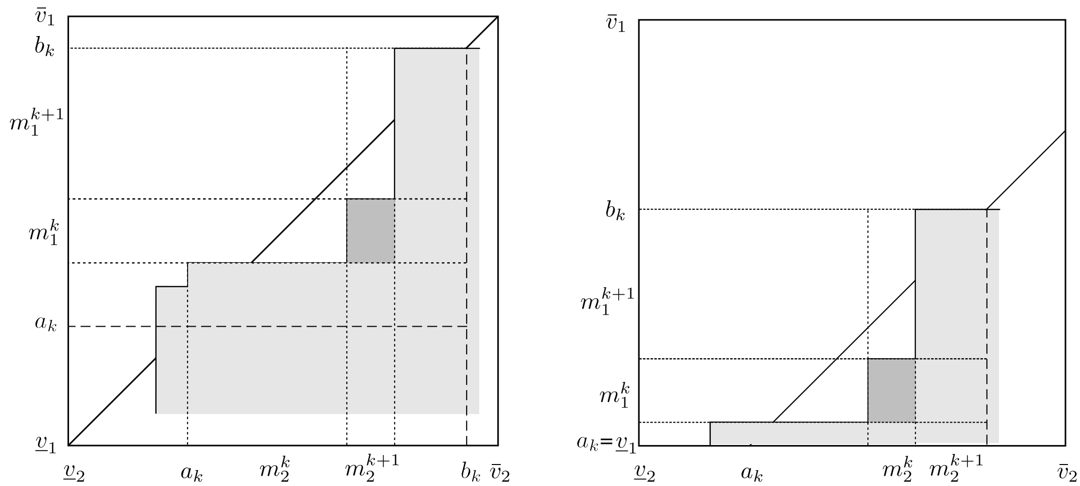
G is budget balanced only if $\psi(Q_G) \leq 0$. According to corollary 2 we have $\psi(Q_G) = \sum_{k_1=1}^{|M_1|} \sum_{k_2=k_1+1}^{|M_2|} \delta\left(m_1^{k_1}, m_2^{k_2}\right) \cdot \left|m_1^{k_1}\right| \cdot \left|m_2^{k_2}\right|$. Denote by K^- the set of all indices for which $\delta\left(m_1^k, m_2^{k+1}\right)$ is negative, i.e., $K^- = \left\{k \mid \delta\left(m_1^k, m_2^{k+1}\right) < 0\right\}$.

Following are two helper lemmas and their proofs. The first implies that the set $\left\{\left(m_1^k, m_2^{k+1}\right)\right\}_{k\in K^-}$ contains *all* the pairs $\left(m_1^{k_1}, m_2^{k_2}\right)$ for which $\delta\left(m_1^{k_1}, m_2^{k_2}\right)$ is negative. The second implies that the indices in K^- cannot be "too close".

Lemma 9 *For every k, if $l > 1$ then $\delta\left(m_1^k, m_2^{k+l}\right) > 0$.*

Proof. Suppose not, so for some k and $l \geq 2$ we have $\delta\left(m_1^k, m_2^{k+l}\right) < 0$. By definition of δ we then have $\bar{m}_1^{k+l-1} < \underline{m}_2^{k+1}$,[13] and thus also $\bar{m}_1^{k+1} < \underline{m}_2^{k+1}$. In that case, however, $E\left[v_2 \mid v_2 \in m_2^{k+1}\right] > E\left[v_1 \mid v_1 \in m_1^{k+1}\right]$, so credibility would imply $p\left(m_1^{k+1}, m_2^{k+1}\right) = 1$, which contradicts non-redundancy under the assumption of $p(m_1^1, m_2^1) = 0$. ■

[13] Note that if m_2^{k+l} exists for some k and l so does m_1^{k+l-1}, because $0 \leq |M_2| - |M_1| \leq 1$



(a) *Example 1* **(b)** *Example 2*

Figure 7: *Budget-balance calculations: the regular case*

Lemma 10 *if* $k \in K^-$ *then* $(k \pm 1) \notin K^-$

Proof. Suppose not and that for some k we have $k+1 \in K^-$, that is $\delta\left(m_1^{k+1}, m_2^{k+2}\right) < 0$. Then by definition $\bar{m}_1^k - \underline{m}_2^{k+1} < 0$ and $\bar{m}_1^{k+1} - \underline{m}_2^{k+2} < 0$. Sum the two inequalities and replace $\bar{m}_1^k$ with $\underline{m}_1^{k+1}$ and $\underline{m}_2^{k+2}$ with $\bar{m}_2^{k+1}$ (which are the same by definition) to get $\frac{\underline{m}_1^{k+1}+\bar{m}_1^{k+1}}{2} < \frac{\bar{m}_2^{k+1}+\underline{m}_2^{k+1}}{2}$ and so $E\left[v_2 \mid v_2 \in m_2^{k+1}\right] > E\left[v_1 \mid v_1 \in m_1^{k+1}\right]$ in contradiction to $p\left(m_1^{k+1}, m_2^{k+1}\right) = 0$, which is implied by non-redundancy and the assumption of $p\left(m_1^1, m_2^1\right) = 0$. A similar argument shows that $(k-1) \notin K^-$. ■

For every $k \in K^-$ define $a_k = \max\left(\underline{m}_2^k, \underline{v}_1\right)$ and $b_k = \min\left(\bar{m}_1^{k+1}, \bar{v}_2\right)$. Figure 7a provides an example for the case of $a_k = \underline{m}_2^k$ and $b_k = \bar{m}_1^{k+1}$. Figure 7b provides an example for the case of $k = 2 \in K^-$ and $a_k = \underline{v}_1$ and $b_k = \bar{m}_1^{k+1}$.

We distinguish between two cases - the regular case, in which for every $k \in K^-$ *either* $\bar{v}_2 > \bar{m}_1^{k+1}$ *or* $\underline{v}_1 < \underline{m}_2^k$, and the boundary case in which there exists $k \in K^-$ for which neither holds.

Regular Case: Denote $\phi(k) = \int_{\underline{m}_2^{k+1}}^{b_k} \int_{a_k}^{\bar{m}_1^k} 1_{(v_1,v_2)\in Q} \cdot \Delta\left(Q_G, v_1, v_2\right) dv_1 dv_2$. Since $b_k \leq \underline{m}_2^{l+1}$

for every $k, l \in K^-$ such that $l > k$,[14] then $\psi(Q_G) > \sum_{k \in K^-} \phi(k)$. For each $k \in K^-$ we have:

$$\begin{aligned} \phi(k) \quad > \quad & \int_{\underline{m}_2^{k+1}}^{\bar{m}_2^{k+1}} \int_{\underline{m}_1^k}^{\bar{m}_1^k} \Delta(Q_G, v_1, v_2)\, dv_1 dv_2 + \int_{\underline{m}_2^{k+2}}^{b_k} \int_{\underline{m}_1^k}^{\bar{m}_1^k} \Delta(Q_G, v_1, v_2)\, dv_1 dv_2 \\ & + \int_{\underline{m}_2^{k+1}}^{\bar{m}_2^{k+1}} \int_{a_k}^{\underline{m}_1^k} \cdot \Delta(Q_G, v_1, v_2)\, dv_1 dv_2 \end{aligned}$$

By definition, $\Delta(Q_G, v_1, v_2) = \delta\left(m_1^k, m_2^{k+1}\right)$ for every $(v_1, v_2) \in \left[\underline{m}_1^k, \bar{m}_1^k\right] \times \left[\underline{m}_2^{k+1}, \bar{m}_2^{k+1}\right]$. Also, since $\Delta(Q_G, v_1, v_2)$ is weakly decreasing in v_1 and weakly increasing in v_2 then $\Delta(Q_G, v_1, v_2) \geq \delta\left(m_1^k, m_2^{k+2}\right)$ for every $(v_1, v_2) \in \left[\underline{m}_1^k, \bar{m}_1^k\right] \times \left[\underline{m}_2^{k+2}, b_k\right]$ and $\Delta(Q_G, v_1, v_2) \geq \delta\left(m_1^{k-1}, m_2^{k+1}\right)$ for every $(v_1, v_2) \in \left[a_k, \underline{m}_1^k\right] \times \left[\underline{m}_2^{k+1}, \bar{m}_2^{k+1}\right]$. Therefore:

$$\begin{aligned} \phi(k) \quad > \quad & \delta\left(m_1^k, m_2^{k+1}\right) \cdot \left|m_1^k\right| \cdot \left|m_2^{k+1}\right| \\ & + \delta\left(m_1^k, m_2^{k+2}\right) \cdot \left|m_1^k\right| \cdot \left(b_k - \underline{m}_2^{k+2}\right) \\ & + \delta\left(m_1^{k-1}, m_2^{k+1}\right) \cdot \left|m_2^{k+1}\right| \cdot \left(\underline{m}_1^k - a_k\right) \end{aligned} \tag{2}$$

Suppose $b_k = \bar{m}_1^{k+1}$ $(\geq \bar{v}_2)$. The third summand in the RHS is non-negative. Since $\bar{m}_1^{k+1} > \bar{m}_2^{k+1}$ then:

$$\phi(k) > \left|m_1^k\right| \cdot \left|m_2^{k+1}\right| \cdot \left(\left(\bar{m}_1^k - \underline{m}_2^{k+1}\right) + \left(\bar{m}_1^{k+1} - \underline{m}_2^{k+2}\right)\right)$$

Note, however, that credibility and $p\left(m_1^{k+1}, m_2^{k+1}\right) = 0$ imply that $\frac{\underline{m}_1^{k+1} + \bar{m}_1^{k+1}}{2} \geq \frac{\underline{m}_2^{k+1} + \bar{m}_2^{k+1}}{2}$. Rearrange to get $\left(\bar{m}_1^k - \underline{m}_2^{k+1}\right) + \left(\bar{m}_1^{k+1} - \underline{m}_2^{k+2}\right) \geq 0$ and thus $\phi(k) > 0$.

If $b_k < \bar{m}_1^{k+1}$ then $a_k = \underline{m}_2^k$ (otherwise this is a boundary case). An analogue argument of omitting the second argument in the RHS of (2) shows that $\phi(k) > 0$. Thus, $\phi(k) > 0$ for every $k \in K^-$, and therefore $\psi(Q_G) > 0$. This is a contradiction to the budget balance of G.

Boundary case: In the boundary case $\bar{v}_2 < \bar{m}_1^{k+1}$ and $\underline{v}_1 > \underline{m}_2^k$ for every $k \in K^-$. We show that G is not optimal by showing that there exists a SPM that implements a posted price and generates more surplus than G.

Note that if $\underline{v}_1 > \underline{m}_2^k$ then $k \leq 2$ (otherwise m_1^{k-2} and m_2^{k-1} are defined, and since $\bar{m}_2^{k-1} = \underline{m}_2^k < \underline{v}_1$ then $\frac{\underline{m}_2^{k-1} + \bar{m}_2^{k-1}}{2} < \frac{\underline{m}_1^{k-2} + \bar{m}_1^{k-2}}{2}$, which contradicts credibility). Also, if $\bar{m}_1^{k+1} > \bar{v}_2$ then $k \geq |M_2| - 2$ (otherwise m_1^{k+2} and m_2^{k+3} are defined, and since $\underline{m}_1^{k+2} = \bar{m}_1^{k+1} > \bar{v}_2$ then $\frac{\underline{m}_2^{k+3} + \bar{m}_2^{k+3}}{2} < \frac{\underline{m}_1^{k+2} + \bar{m}_1^{k+2}}{2}$, which contradicts credibility).

The fact that $\bar{v}_2 < \bar{m}_1^{k+1}$ and $\underline{v}_1 > \underline{m}_2^k$ for every $k \in K^-$ implies that either $\big(|M_2| = 3$ and

[14] To see why $b_k \leq \underline{m}_2^{l+1}$ when $l > k$, note that according to lemma 10 it must be that $l \geq k+2$. If $l > k+2$ then $\bar{m}_1^{k+1} \leq \underline{m}_2^{k+4}$, or otherwise there is a contradiction to $p\left(m_1^{k+2}, m_2^{k+3}\right) = 1$. If $l = k+2$ then $l \in K^-$ implies $b_k < \underline{m}_2^{k+3}$.

$\bar{v}_2 < \bar{m}_1^2$) or ($|M_2| = 3$ and $\underline{v}_1 > \bar{m}_2^2$) or ($|M_2| = 4$ and $\underline{v}_1 > \bar{m}_2^2$ and $\bar{v}_2 < \bar{m}_1^3$)). Assume that the latter holds, as illustrated in figure 8.[15] For convenience we normalize $\underline{m}_1^1 = 0, \bar{m}_2^4 = 1$ and denote $\bar{m}_1^1 \equiv \beta$, $\bar{m}_2^2 \equiv x$, $\bar{m}_1^2 \equiv y$, $\bar{m}_2^3 \equiv 1-\varepsilon$. Since $\delta\left(m_1^2, m_2^3\right) < 0$ then $x > y$.

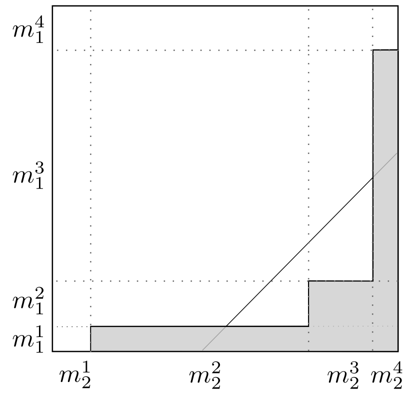


Figure 8: *Budget-balance calculations: the boundary case.*

Optimality of G implies $E\left[v_2 \mid v_2 \in m_2^3\right] = E\left[v_1 \mid v_1 \in m_1^3\right]$. Otherwise, if $E\left[v_2 \mid v_2 \in m_2^3\right] < E\left[v_1 \mid v_1 \in m_1^3\right]$ then decreasing $\underline{m}_1^4$ by a small amount would increase the surplus, decrease the deficit and would not affect credibility. For the same reason it must be that $E\left[v_2 \mid v_2 \in m_2^2\right] = E\left[v_1 \mid v_1 \in m_1^2\right]$. We therefore have $\bar{m}_2^1 = (\beta + y - x)$ and $\bar{m}_1^3 = (1-\varepsilon) + x - y$.

Optimality of G implies also that either $\varepsilon = 0$ or $\beta = 0$. Writing the surplus and deficit generated by G explicitly yields:

$$\begin{aligned} D(G) &= (-x)\varepsilon^2 + x(x-2y+1)\varepsilon + (y-1)\beta^2 - (y-1)(2x-y)\beta + y(x-1)(x-y) \\ S(G) &= -\frac{1}{2}(1-x)\varepsilon^2 + \frac{1}{2}(1-x)(x-2y+1)\varepsilon - \frac{1}{2}y(x-\beta-1)(x-y-\beta+1) \end{aligned}$$

Since G is optimal then $D(G) = 0$.[16] Suppose that $x + y > 1$. Solving for $\beta(\varepsilon)$ such that $D(G) = 0$, and plugging it into $S(G)$ gives the surplus as a function of ε, s.t. $D(G) = 0$. Taking the first derivative gives:

$$\frac{dS}{d\varepsilon}|_{D(G)=0} = \frac{1}{(1-y)}(1-x-y)\left(\frac{x-2y+1}{2} - \varepsilon\right)$$

Note that $\varepsilon = \frac{x-2y+1}{2}$ implies $D(G) > 0$, and since $\frac{\partial D(G)}{\partial \varepsilon} > 0$, it must be that $\varepsilon < \frac{x-2y+1}{2}$ which implies $\frac{dS}{d\varepsilon}|_{D(G)=0} < 0$. Thus $\varepsilon = 0$ maximizes $S|_{D(G)=0}$ and generates the suprlus $S_{\varepsilon=0} = \frac{1}{2}y(1-x)\left(\frac{x-2y+1}{1-y}\right)$. Note, however, that the surplus generated by setting a posted price y is $S_{pp\_y} = \frac{1}{2}y(1-y)$. The difference $S_{pp\_y} - S_{\varepsilon=0} = \frac{1}{2}\frac{y}{1-y}(x-y)^2$ is positive, and this is a contra-

[15] For the other two cases the proof is similar.

[16] Otherwise, for example, increasing y would increase the surplus without violating credibility.

diction to the optimality of G.

If $x+y<1$ the proof is similar, only that in this case $\beta=0$ is optimal and it generates less surplus than the posted price x.

9.6 Proof of Lemma 7

We start by constructing a quasi-monotone cover $\mathcal{Q}^*(G)$ for Q_G. Denote by Ω set of all environments at period $T-1$ of all histories of length T in G, that is $\Omega=\{\mathbf{m}_{T-1}\mid(\mathbf{m}_1,..,\mathbf{m}_{T-1},\mathbf{m}_T)\in H\}$, and denote by ω an element in Ω. The set Ω is the set of all the environments in which a posted price is going to be implemented at period T. Therefore, for all $\omega\in\Omega$, the set $Q_G\cap\omega$ (which is the set of pairs that trade at T) is quasi monotone.

Let $\bar{\omega}_i$ and $\underline{\omega}_i$ denote the upper and lower bounds of ω_i, respectively. Since $\omega\in\Omega$ is an environment then $\bar{\omega}_1,\underline{\omega}_1,\bar{\omega}_2$ and $\underline{\omega}_2$ are all well defined. $\mathcal{Q}^*(G)$ is constructed by "attaching" to every $\omega\in\Omega$ a subset of Q_G that contains pairs $(v_1,v_2)\in Q_G$ such that $v_1<\bar{\omega}_1$ and $v_2>\underline{\omega}_2$. This is done such that every pair $(v_1,v_2)\in Q_G$ is attached to exactly one ω.

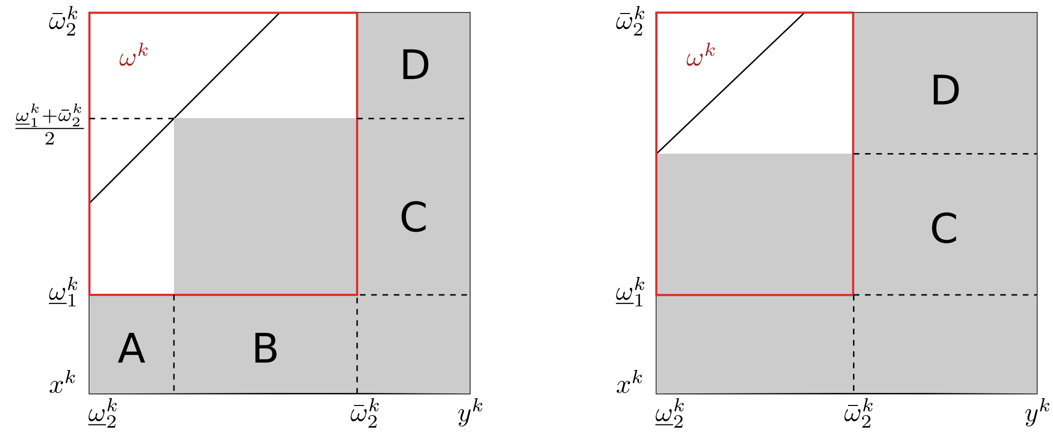
Let us order all the elements $\omega\in\Omega$ in an ascending, with respect to $\underline{\omega}_2$, and if $\underline{\omega}_2$ is equal for two elements then according to $\underline{\omega}_1$. Denote by ω^k the k^{th} element in Ω according to this order. For convenience denote $\underline{\omega}_2^{|\Omega|+1}\equiv\bar{v}_2$. For each ω^k, define $r(k)$ to be the lowest element l such that $l>k$ and $\underline{\omega}_1^l<\bar{\omega}_1^k$. If such l does not exists we define $r(k)$ to be $|\Omega|+1$. Formally:

$$r(k)=\begin{cases}|\Omega|+1 & \left\{\omega^l\;\mid\;\underline{\omega}_1^l<\bar{\omega}_1^k\text{ and }l>k\right\}=\varnothing\\ \min\left\{l\;\mid\;\underline{\omega}_1^l<\bar{\omega}_1^k\text{ and }l>k\right\} & \text{otherwise}\end{cases}$$

We attach subsets of Q_G to the elements in Ω in an iterative process, according to the order defined above. The set of pairs that is attached to ω^k is denoted Q_k, and defined as follows – it contains all the pairs $(v_1,v_2)\in Q$ such that $\underline{v}_1\le v_1<\underline{\omega}_1^k$ and $\underline{\omega}_2^k<v_2<\underline{\omega}_2^{r(k)}$ *that have not already been attached* to Q_j with $j<k$. Formally $Q_k=\left\{Q/\cup_{j=1}^{k-1}Q_j\right\}\cap\left(\left[\underline{v}_1,\bar{\omega}_1^k\right]\times\left[\underline{\omega}_2^k,\underline{\omega}_2^{r(k)}\right]\right)$. Figure 6b in the text shows sets Q_k that were formed according to this process.

By construction it is clear that every $(v_1,v_2)\in Q_G$ is contained by no more than one set Q_k. Since every $(v_1,v_2)\in Q_G$ such that $\bar{\omega}_1^k<v_1<\underline{\omega}_1^k$ for some k is contained in either Q_k or $Q_{r(k)}$, then each (v_1,v_2) is contained in exactly one Q_k. Also, since $Q_G\cap\omega_k$ is quasi-monotone for every k, then every Q_k is also quasi-monotone. Therefore $\mathcal{Q}^*(G)\equiv\{Q_k\}_{k=1}^{|\Omega|}$ is a quasi monotone cover of G.

We continue by showing that $\psi(Q_k)\ge 0$ for every $Q_k\in\mathcal{Q}^*(G)$. Denote $x^k=\min\{v_1\mid(v_1,v_2)\in Q_k\}$ and $y^k=\max\{v_2\mid(v_1,v_2)\in Q_k\}$. Recall that for every k, the optimal posted price is implemented



(a) *Case 1* **(b)** *Case 2*

Figure 9: *Representative elements in* $\mathcal{Q}^*(G)$

in the environment $\omega^k \in \Omega$ at period T.

Suppose that $\underline{\omega}_2^k < \frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2} < \bar{\omega}_1^k$, as illustrated in figure 9a. Consider the area A in the figure, and denote by $\Delta(A)$ be the value of $\Delta(Q_k, v_1, v_2)$ of a representative pair (v_1, v_2) in A (see section 5 for the definition of Δ). It is easy to see that $\Delta(A) = \left(\underline{\omega}_1^k - \underline{\omega}_2^k\right)$. Similarly $\Delta(B) = \left(\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2} - \underline{\omega}_2^k\right), \Delta(C) = \left(\bar{\omega}_1^k - \frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}\right)$ and $\Delta(D) = \left(\bar{\omega}_1^k - \bar{\omega}_2^k\right)$. Therefore:

$$\begin{aligned}
\psi(Q_k) &\geq \int_{x^k}^{\underline{\omega}_1^k} \int_{\underline{\omega}_2^k}^{\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}} \Delta(A)\, dv_2 dv_1 + \int_{x^k}^{\underline{\omega}_1^k} \int_{\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}}^{\bar{\omega}_2^k} \Delta(B)\, dv_2 dv_1 \\
&\quad + \int_{\underline{\omega}_1^k}^{\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}} \int_{\bar{\omega}_2^k}^{y^k} \Delta(C)\, dv_2 dv_1 + \int_{\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}}^{\bar{\omega}_1^k} \int_{\bar{\omega}_2^k}^{y^k} \Delta(D)\, dv_2 dv_1 \\
&= \left(\underline{\omega}_1^k - x^k\right)\left(\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2} - \underline{\omega}_2^k\right)^2 + \left(y^k - \bar{\omega}_2^k\right)\left(\bar{\omega}_1^k - \frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}\right)^2 \\
&\geq 0
\end{aligned}$$

and the inequality is strict if either $\left(\underline{\omega}_1^k > x^k\right)$ or $\left(y^k > \bar{\omega}_2^k\right)$.

Suppose on the other hand that $\frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2} < \underline{\omega}_2^k$ as illustrated in figure 9b. Let $\Delta(C)$and $\Delta(D)$ be defined as above, and now $\Delta(C) = \left(\bar{\omega}_1^k - \underline{\omega}_2^k\right)$ and $\Delta(D) = \left(\bar{\omega}_1^k - \bar{\omega}_2^k\right)$. Therefore

$$\begin{aligned}
\psi\left(Q^k\right) &\geq \int_{\underline{\omega}_1^k}^{\underline{\omega}_2^k} \int_{\bar{\omega}_2^k}^{y^k} \Delta(C)\, dv_2 dv_1 + \int_{\underline{\omega}_2^k}^{\bar{\omega}_1^k} \int_{\bar{\omega}_2^k}^{y^k} \Delta(D)\, dv_2 dv_1 \\
&= \left(y^k - \bar{\omega}_2^k\right)\left(\bar{\omega}_1^k - \underline{\omega}_2^k\right)\left(\left(\underline{\omega}_2^k - \underline{\omega}_1^k\right) + \left(\bar{\omega}_1^k - \bar{\omega}_2^k\right)\right) \\
&> \left(y^k - \bar{\omega}_2^k\right)\left(\bar{\omega}_1^k - \underline{\omega}_2^k\right)\left(\bar{\omega}_1^k - \frac{\underline{\omega}_1^k+\bar{\omega}_2^k}{2}\right) > 0
\end{aligned}$$

and the inequality is strict if $y^k > \bar{\omega}_2^k$. The case in which $\bar{\omega}_1^k < \frac{\underline{\omega}_1^k + \bar{\omega}_2^k}{2}$ is analogue. Thus $\psi\left(Q^k\right) \geq 0$ for every $Q_k \in \mathcal{Q}^*(G)$.

Furthermore, if Ω is not a singleton (which is the case when there exist $h, h' \in H/Z$ such that $L(h) = L(h')$) then at least for one element ω^k, either $\underline{\omega}_1^k > \min\{v_1 \mid (v_1, v_2) \in Q_k\}$ or $\bar{\omega}_2^k < \max\{v_2 \mid (v_1, v_2) \in Q_k\}$ and thus $\psi(Q_k) > 0$ for at least for one element in $Q_k \in \mathcal{Q}^*(G)$.

9.7 Proof of Lemma 8

The proof begins with definitions of some useful functions, continues with two lemmas that highlight characteristics of these functions, and concludes by applying the lemmas to the case of a SPM with a DMT that has a quasi-monotone cover.

For every environment $F = (F_1 \times F_2) \subseteq V$ and a DMT $Q \subseteq F$ let:

$$\begin{aligned}
\bar{q}_i(Q, F, v_i) &= \int_{v_{-i} \in F_{-i}} 1_{(v_1, v_2) \in Q} \cdot \left(\frac{1}{\bar{F}_{-i} - \underline{F}_{-i}}\right) dv_{-i} \\
R_1(Q, F) &= \int_{v_1 \in F_1} \bar{q}_1(Q, F, v_1) \cdot \left(\frac{v_1 - \underline{F}_1}{\bar{F}_1 - \underline{F}_1}\right) dv_1 \\
R_2(Q, F) &= \int_{v_2 \in F_2} \bar{q}_2(Q, F, v_2) \cdot \left(1 - \frac{v_2 - \underline{F}_2}{\bar{F}_2 - \underline{F}_2}\right) dv_2 \\
S(Q, F) &= \int_{v_1 \in F_1} \int_{v_2 \in F_2} \left((v_2 - v_1) \cdot 1_{(v_i, v_{-i}) \in Q}\right) \cdot \left(\frac{1}{\bar{F}_1 - \underline{F}_1}\right) \left(\frac{1}{\bar{F}_2 - \underline{F}_2}\right) dv_2 dv_1
\end{aligned}$$

where $\bar{F}_i$ denotes the highest type in F_i and $\underline{F}_i$ denotes the lowest one. Let $F_i^B(Q) = \{v_i \mid (v_1, v_2) \in Q \text{ for some } v_{-i} \in V_{-i}\}$ and $F^B(Q) = F_1^B(Q) \times F_2^B(Q)$. We refer to $F^B(Q)$ as the *bounding environment* of Q, this is the smallest environment that contains all the elements in Q. Finally, let $\xi(Q) = \left(\frac{\bar{F}_1^B(Q) - \underline{F}_1^B(Q)}{\bar{v}_1 - \underline{v}_1}\right) \cdot \left(\frac{\bar{F}_2^B(Q) - \underline{F}_2^B(Q)}{\bar{v}_2 - \underline{v}_2}\right)$, this is the ex ante expected probability that the buyer's type is in $F_2^B(Q)$ and the seller's type is in $F_1^B(Q)$.

If Q is induced by a mechanism which is incentive compatible in the environment F, then the quantities $R_i(Q, F)$ and $S(Q, F)$ are the information rent that has to be paid to agent i and the surplus generated by the mechanism, respectively. If, in addition, the lowest type of the buyer and the highest type of the seller in F expect payoff 0, then the deficit of the mechanism is given by $R_1(Q, F) + R_1(Q, F) - S(Q, F)$. The functions $R_i(Q, F)$ and $S(Q, F)$ are defined, however, even if Q is not induced by an incentive compatible mechanism.

We continue by stating and proving two lemmas:

Lemma 11 *If Q is a DMT and Q' is a submap of Q then $R_i(Q, V) = R_i(Q', V) + R_i(Q \backslash Q', V)$*

and $S(Q,V) = S(Q',V) + S(Q\backslash Q',V)$.

Proof. We show the explicit proof for the case of $i=2$. The proof for $i=1$ is similar and is omitted. Note that $\bar{q}_i(Q,V,v_i) = \bar{q}_i(Q',V,v_i) + \bar{q}_i(Q\backslash Q',V,v_i)$ for every $v_i \in V_i$. Therefore:

$$\begin{aligned} R_2(Q,V) &= \int_{v_2 \notin F_2(Q')} \bar{q}_2(Q,V,v_2)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \\ &\quad + \int_{v_2 \in F_2(Q')} \bar{q}_2\left(Q\backslash Q',V,v_2\right)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \\ &\quad + \int_{v_2 \in F_2(Q')} \bar{q}_2\left(Q',V,v_2\right)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \end{aligned}$$

Since $\bar{q}_2(Q',V,v_2) = 0$ for $v_2 \notin F_2(Q')$ then the third element of the RHS is in fact $R_2(Q',V)$, and since $\bar{q}_2(Q,V,v_2) = \bar{q}_2(Q\backslash Q',V,v_2)$ for $v_2 \notin F_2(Q')$ then:

$$\begin{aligned} R_2(Q,V) &= \int_{v_2 \in V_2} \bar{q}_2\left(Q\backslash Q',V,v_2\right)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 + R_2\left(Q',V\right) \\ &= R_2\left(Q\backslash Q',V\right) + R_2\left(Q',V\right) \end{aligned}$$

The equality $S(Q,V) = S(Q,V) + S(Q\backslash Q,V)$ follows directly from $1_{(v_i,v_{-i})\in Q} = 1_{(v_i,v_{-i})\in Q'} + 1_{(v_i,v_{-i})\in Q\backslash Q'}$. ■

Lemma 12 $R_i(Q,V) \geq \xi(Q) \cdot R_i\left(Q,F^B(Q)\right)$ *and* $S(Q,V) = \xi(Q) \cdot S\left(Q,F^B(Q)\right)$

Proof. We show the explicit proof for the case of R_2. The proof for R_1 is similar and is omitted. Since $\bar{q}_2(Q,V,v_2) = 0$ for $v_2 \notin F_2^B(Q)$ then:

$$\begin{aligned} R_2(Q,V) &= \int_{v_2 \in V_2} \bar{q}_2(Q,V,v_2)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \\ &= \int_{v_2 \in F_2^B(Q)} \bar{q}_2(Q,V,v_2)\left(1 - \frac{v_2 - \underline{v}_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \\ &\geq \int_{v_2 \in F_2^B(Q)} \bar{q}_2(Q,V,v_2)\left(\frac{\bar{F}_2^B(Q) - v_2}{\bar{v}_2 - \underline{v}_2}\right) dv_2 \end{aligned}$$

Since $\bar{q}_2(Q,V,v_2) = \frac{\bar{F}_1^B(Q) - \underline{F}_1^B(Q)}{\bar{v}_1 - \underline{v}_1} \cdot \bar{q}_2\left(Q,F^B(Q),v_2\right)$, then:

$$\begin{aligned} &= \int_{v_2 \in F_2^B(Q)} \bar{q}_2\left(Q,F^B(Q),v_2\right)\left(\frac{\bar{F}_2^B(Q) - v_2}{\bar{v}_2 - \underline{v}_2}\right) \cdot \left(\frac{\bar{F}_1^B(Q) - \underline{F}_1^B(Q)}{\bar{v}_1 - \underline{v}_1}\right) \cdot dv_2 \\ &= \xi(Q) \cdot \int_{v_2 \in F_2^B(Q)} \left[\bar{q}_2\left(Q,F^B(Q),v_2\right) \cdot \left(1 - \frac{v_2 - \underline{F}_2(Q)}{\bar{F}_2^B(Q) - \underline{F}_2^B(Q)}\right)\right] dv_2 \\ &= \xi(Q) \cdot R_2\left(Q,F^B(Q)\right) \end{aligned}$$

To show that $S(Q,V) = \xi(Q) \cdot S\left(Q, F^B(Q)\right)$, write:

$$\begin{aligned} S(Q,V) &= \int_{v_1 \in V_1} \int_{v_2 \in V_2} (v_2 - v_1) \cdot 1_{(v_i, v_{-i}) \in Q} \cdot \frac{1}{\bar{v}_1 - \underline{v}_1} \frac{1}{\bar{v}_2 - \underline{v}_2} dv_2 dv_1 \\ &= \xi(Q) \cdot \int_{v_1 \in V_1} \int_{v_2 \in V_2} \left((v_2 - v_1) \cdot 1_{(v_i, v_{-i}) \in Q}\right) \\ &\qquad \cdot \frac{1}{\bar{F}_1^B(Q) - \underline{F}_1^B(Q)} \frac{1}{\bar{F}_2^B(Q) - \underline{F}_1^B(Q)} dv_2 dv_1 \\ &= \xi(Q) \cdot S\left(Q, F^B(Q)\right) \end{aligned}$$

■

Suppose now that G is a mediator in the environment V and that $\mathcal{Q}$ is a quasi-monotone cover of Q_G. It follows from lemma 11 that $R_i(Q_G, V) = \sum_{Q_k \in \mathcal{Q}} R_i(Q_k, V)$ and $S(Q_G, V) = \sum_{Q_k \in \mathcal{Q}} S(Q_k, V)$. Since G is incentive compatible then $D(G) = R_1(Q_G, V) + R_2(Q_G, V) - S(Q_G, V)$ which, in turn, implies that $D(G) = \sum_{Q_k \in \mathcal{Q}} R_1(Q_k, V) + R_2(Q_k, V) - S(Q_k, V)$, and by lemma 12 therefore $D(G) \geq \sum_{Q_k \in \mathcal{Q}} R_1\left(Q_k, F^B(Q_k)\right) + R_2\left(Q_k, F^B(Q_k)\right) - S\left(Q_k, F^B(Q_k)\right)$.

For every Q_k denote by Γ^k an incentive compatible mechanism in the environment $F^B(Q_k)$ that induces Q_k, in which the types $\bar{F}_1^B(Q_k)$ and $\underline{F}_2^B(Q_k)$ expect the payoff 0. The deficit of Γ^k is given by $D\left(\Gamma^k\right) = R_1\left(Q_k, F^B(Q_k)\right) + R_2\left(Q_k, F^B(Q_k)\right) - S\left(Q_k, F^B(Q_k)\right)$ on one hand, and on the other hand equals $\psi(Q_k)$ (according to proposition 1). Therefore $D(G) \geq \sum_{Q_k \in \mathcal{Q}} \psi(Q_k)$.

10 References

1. Baliga, S., Corchon, L. and Sjostrom, T. (1997), "The theory of implementation when the planner is a player", Journal of Economic Theory, 77(1), 15-33.

2. Bester, H. and Strausz, R., (2000), "Imperfect commitment and the revelation principle: the multi-agent case", *Economics Letters*, 69(2), 165-171

3. Bester, H. and Strausz, R., (2001), "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case", *Econometrica*, 69(4), 1077-98.

4. Blume A., Board, O.J., and Kawamura, K. (2007): "Noisy talk", *Theoretical Economics*, 2 (4), 395-440.

5. Borgers, T. and Norman P., (2008), "A Note on Budget Balance under Interim Participation Constraints: The Case of Independent Types", *Economic Theory,* 34.

6. Chakravorty, B., Corchon, L. and Wilkie, S. (2006), "Credible implementation", *Games and*

Economic Behavior, 57(1), 18-36.

7. Crawford, V. and Sobel, J., (1982), "Strategic Information Transmission," *Econometrica*, 50(6), 1431-51.
8. Forges, F. (1986), “An approach to communication equilibria.” *Econometrica*, 54,1375–1385.
9. Gibbons, R. (1988), “Learning in Equilibrium Models of Arbitration”, *The American Economic Review*, 78(5), 896-912.
10. Goltsman, M., Horner, J., Pavlov, G. and Squintani, F. (2009), "Mediation, arbitration and negotiation", *Journal of Economic Theory*, 144(4), 1397-1420.
11. Ivanov, M. (2010), "Communication via a strategic mediator", *Journal of Economic Theory*, 145 (2) 869–884.
12. Myerson, R. (1986), “Multistage games with communication.” *Econometrica*, 54,323–358.
13. Myerson, R. and Satterthwaite, M. (1983), "Efficient mechanisms for Bilateral Trading", *Journal of Economic Theory*, 29(2), 265-281.
14. Osborne, M. and Rubinstein, A., (1994), A course in game theory, MIT Press.
15. Sobel, J., (2013). Giving and receiving advice. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics. Cambridge University Press, pp. 305–341, Chap. 10.
16. Skreta, V. (2006), “Sequentially Optimal Mechanisms”, *The Review of Economic Studies*, 73 (4), 1085-1111.
17. Skreta, V. (2013), "Optimal Auction Design Under Non-Commitment" Working Papers 13-08, New York University, Leonard N. Stern School of Business, Department of Economics.
18. Vartiainen, H. (2013), “Mechanism design without commitment”, Working paper.